

КІБЕРНЕТИКА та КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК 519.67

DOI:10.34229/2707-451X.23.4.7

В.В. ТРЕТИНИК, Є.О. БУДЗІНСЬКИЙ

ВИКОРИСТАННЯ МЕТОДІВ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ

Вступ. Художники завжди використовували різні засоби для вираження своєї творчості та дослідження своєї уяви. З розвитком цифрових технологій митці нині мають доступ до величезної кількості інструментів, які дозволяють їм створювати витвори мистецтва, які є більш витонченими, складними та візуально захоплюючими, ніж будь-коли раніше. Останнім часом зростає інтерес до використання штучного інтелекту для створення зображень для художніх цілей.

Створення зображень для сфери мистецтва передбачає застосування алгоритмів і методів машинного навчання (МН) для створення цифрових творів мистецтва [1, 2], які можуть імітувати стилі, техніки й естетику традиційних форм мистецтва. Системи ШІ можуть навчатися на величезних масивах даних, щоб створювати зображення, які є неймовірно реалістичними та деталізованими, а також унікальними та оригінальними.

Мета даної роботи – застосування методів глибокого навчання для генерації зображень творів мистецтва.

На сьогоднішній день одним із найбільш часто використовуваних підходів для створення образів у мистецтві є генеративні змагальні мережі (Generative Adversarial Network, GAN) [3]. GAN – це моделі глибокого навчання, які здатні генерувати нові зображення, які мають схожий набір характеристик до існуючих зображень, що надано в навчальному наборі даних. GAN складаються з двох нейронних мереж: мережі генератора, яка створює нові зображення, і мережі дискримінатора, яка намагається відрізнити згенеровані зображення від реальних.

У даній роботі запропоновано скомбінувати два типи моделей машинного навчання, а саме: архітектуру згорткових нейронних мереж (Convolutional Neural Network, CNN) CNN та модель GAN. Об'єднання такого типу, зумовлене тим що CNN чудово підходять для обробки зображень, а GAN мають необхідний генеративний потенціал. Такий клас моделей носить назву глибока згорткова генеративна змагальна мережа (Deep Convolutional Generative Adversarial Network, DCGAN). В загальному випадку, моделі типу DCGAN навчаються швидше та генерують зображення вищої якості ніж звичайні GAN моделі [4].

Розглянуто застосування методів глибокого навчання для генерації зображень. Проведено порівняльний аналіз існуючих засобів генерації зображень. Запропонована модифікація генеративної моделі. Розроблена система генерує зображення фіксованого розміру (64 × 64, 256 × 256, 1024 × 1024) художнього характеру, мінімальне значення індексу FID, при навчанні – 128.

Ключові слова: згорткові нейронні мережі, машинне навчання, генерація мистецтва, генеративна змагальна мережа.

© В.В. Третиник, Є.О. Будзінський, 2023

Загальна абстрактна схема генеративної моделі GAN, показана на рис. 1.

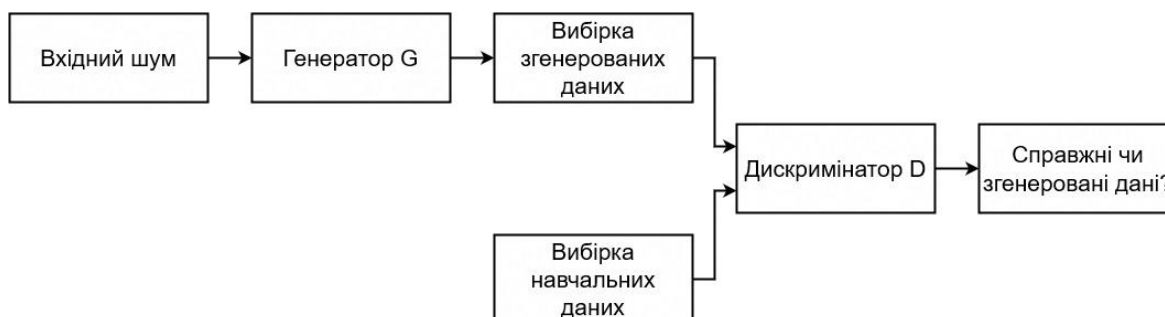


РИС. 1. Абстрактна схема моделі GAN [4]

Розглянемо основні програмні засоби та системи для генерації зображень, що використовують принцип генеративних моделей: Dall-E 2, Stable Diffusion, Photosonic, Imagen.

Як основні характеристики для порівняння готових рішень, було виділено наступні: статус, ціна, стиль генерації, кількість зображень, контроль генерації, обмеження на використання.

Дослідження кожної окремої реалізації було ускладнено, оскільки більшість систем досить нові, та знаходяться у фазі закритого тестування, або не розповсюджують деталі та особливості програмної реалізації конкретної використаної моделі. Незважаючи на це, в результаті аналізу, було сформовано порівняльну таблицю існуючих засобів, яка наведена у табл. 1.

ТАБЛИЦЯ 1. Порівняння існуючих засобів генерації зображень

	Dall-E 2	Stable Diffusion	Photosonic	Imagen
Статус	Закрите тестування	Вільний доступ	Вільний доступ	Закрите тестування
Ціна	Невідомо	Безкоштовно	Частково безкоштовно	Невідомо
Стиль генерації	Довільний	Довільний	Широкий набір	Довільний
Кількість зображень	Теоретично довільна	До 4	Теоретично довільна	Теоретично довільна
Контроль генерації	Текст	Текст	Текст, стиль	Текст
Обмеження на використання	Невідомо	Демо-версія	Демо-версія (5 генерацій)	Невідомо

Наведений список засобів генерації зображень, у загальному випадку, вирішують більш широкий клас завдань, а саме: генерацію зображень із запропонованого тексту. Але при цьому мають багато обмежень, а саме: можливість вільного використання та обмеження кількості згенерованих зображень за один раз.

Мета розробки нашої системи це вирішення проблеми доступності систем генерації зображень та розширення кількості зображень, що можна отримати за один процес генерації. Але додатково

накладаючи обмеження на стиль та вид зображень, що буде згенеровано і також обмежуючи контроль над генерацією зображень. Іншими словами, система буде генерувати зображення тих самих стилів, що було наведено у навчальному наборі даних (10 художніх стилів). Кількість зображень, що можна згенерувати за один раз, буде обмежено до 100 одиниць. А контроль над генерацією буде майже відсутнім.

Для динамічного представлення системи генерації зображень було створено діаграму основних компонентів, що показана на рис. 2.

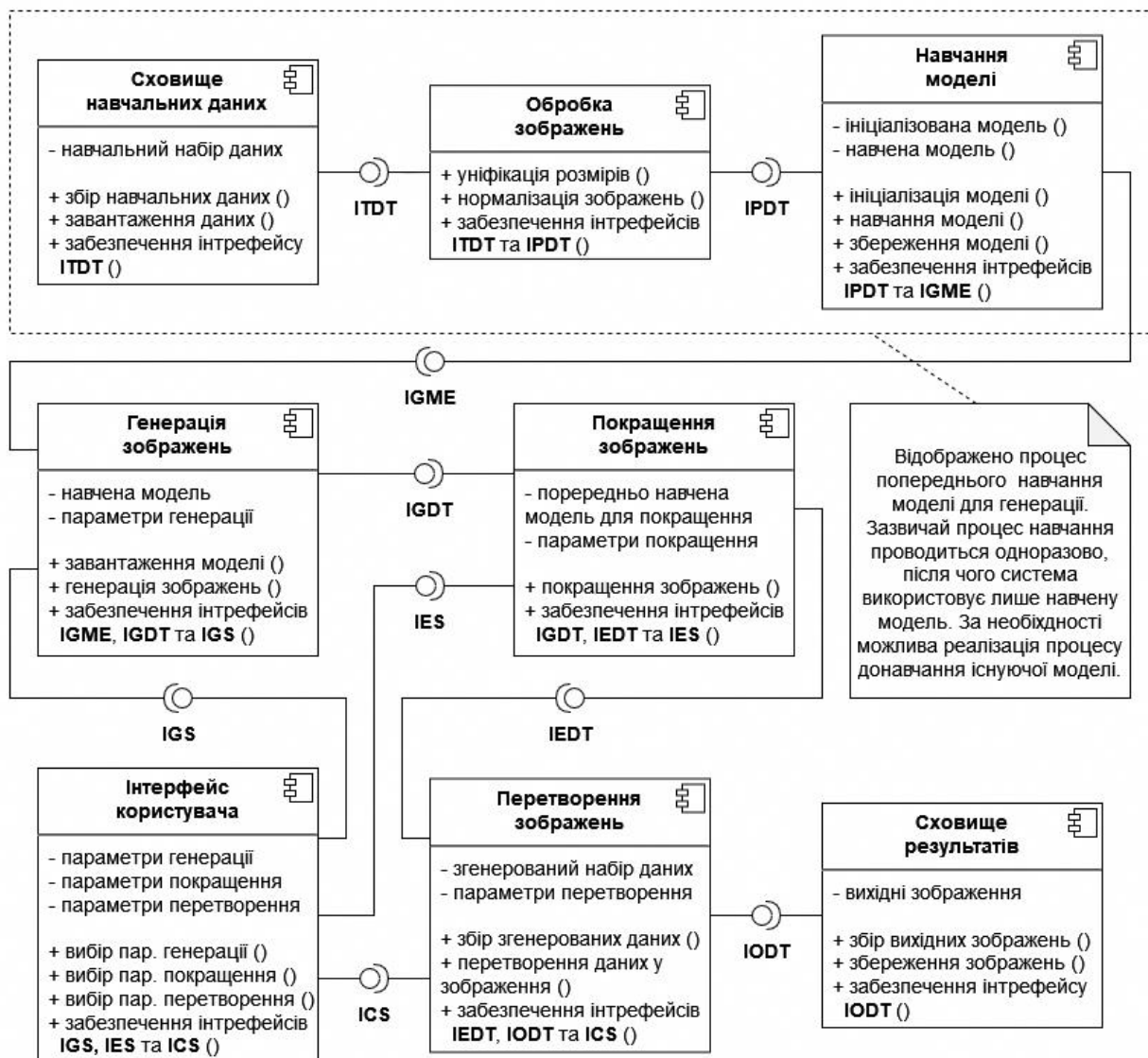


РИС. 2. Модель системи генерації зображень. Діаграма компонентів у нотатції UML

Опишемо функціональність і призначення компонентів системи:

- *сховище навчальних даних*: локальне сховище даних, на якому розміщено зібраний набір навчальних зображень, що буде використано для навчання відповідної моделі генерації зображень;

- *обробка зображень*: набір інструментів, що проводять попередню обробку навчальних зображень для їх подальшого застосування у процесі навчання моделі генерації зображень;
- *навчання моделі*: набір інструментів, що дозволяє створити модель необхідної структури та провести її навчання на відповідному навчальному наборі даних;
- *генерація зображень*: набір інструментів, що використовуючи навчену генеративну модель та реалізує процес створення необхідних зображень із заданими параметрами;
- *покращення зображень*: попередньо навчена модель на основі існуючого методу Real-ESRGAN [5], що може використовуватись для збільшення розмірів згенерованих зображень;
- *перетворення зображень*: набір інструментів, що реалізує перетворення даних, отриманих у процесі функціонування системи, що репрезентують зображення, зворотно у формат зображень;
- *інтерфейс користувача*: компонент, який уособлює елементарний графічний інтерфейс користувача для збору відповідних параметрів генерації зображень та відображення процесу функціонування системи в цілому (відображення прогресу генерації зображень, поточний стан системи, обраний набір параметрів генерації тощо);
- *сховище результатів*: локальне сховище даних, на якому буде розміщено результуючі згенеровані зображення, що було отримано у процесі роботи системи генерації зображень.

Список інтерфейсів:

- **ITDT** (Interface of Training Data Transferring) – інтерфейс передавання навчальних даних (із модуля сховища навчальних даних) на вхід модуля *обробки зображень*;
- **IPDT** (Interface of Processed Data Transferring) – інтерфейс передавання оброблених навчальних даних (із модуля обробки зображень) на вхід модуля *навчання моделі*;
- **IGME** (Interface of Generative Model Extraction) – інтерфейс отримання результуючої навченої генеративної моделі (із модуля Навчання моделі) на вхід модуля *генерації зображень*;
- **IGDT** (Interface of Generated Data Transferring) – інтерфейс передавання згенерованих даних (із модуля Генерації зображень) на вхід модуля *покращення зображень*;
- **IEDT** (Interface of Enhance Data Transferring) – інтерфейс передавання згенерованих зображень більшого розміру (із модуля Покращення зображень) на вхід модуля *перетворення зображень*;
- **IODT** (Interface of Output Data Transferring) – інтерфейс передавання даних у форматі зображень (із модуля Перетворення зображень) на вхід модуля *сховище результатів*;
- **IGS** (Interface of Generating Settings) – інтерфейс передавання параметрів генерації (із модуля Інтерфейсу користувача) на вхід модуля *генерації зображень*;
- **IES** (Interface of Enhancing Settings) – інтерфейс передання параметрів покращення (із модуля Інтерфейсу користувача) на вхід модуля *покращення зображень*;
- **ICS** (Interface of Converting Settings) – інтерфейс передання параметрів перетворення (із модуля Інтерфейсу користувача) на вхід модуля *перетворення зображень*.

У даній роботі як вхідні дані було обрано датасет «ArtBench-10» [6]. Він містить 60 000 зображень творів мистецтва за 10-ма відмінними художніми стилями (модерн, бароко, експресіонізм, імпресіонізм, постмодернізм, реалізм, ренесанс, романтизм, сюрреалізм, укійо-е). Кожен художній стиль містить по 6 000 зображень. Загальні розміри зображень є довільними. Кожне зображення було приведено до квадратного вигляду, відкиданням зайвих смуг пікселів по горизонталі (чи вертикалі) з обох боків, центруючи результуюче зображення. Загальний розмір зображень також було уніфіковано до 64x64 пікселі, використовуючи бікубічний алгоритм зменшення. Для подальшого використання даного набору, кожне зображення було представлено у вигляді 3 кольорових каналів (червоний, зелений, синій), із 8 бітним значеннями інтенсивності кожен (від 0 до 255). Після чого, кожний кольоровий канал, використовуючи лінійні перетворення, було приведено до відрізка $[-1; 1]$, для покращення ефективності навчання моделі.

В процесі попереднього навчання та пошуку модифікації моделей типу GAN, було реалізовано та проведено навчання, щонайменше, 14 модифікацій. Для аналізу було використано три модифікації, для яких вдалось отримати задовільні результати, а саме: «DCGAN», «DCGAN-WI» та «WGAN-GP» (наведені назви модифікацій є арбітральними, та відображають загальну ідею модифікації моделі, і не є загальноприйнятими). Під час пошуку оптимальної структури модифікації, було розглянуто певний перелік значень можливих гіперпараметрів відповідної генеративної моделі типу GAN (табл. 2).

ТАБЛИЦЯ 2. Гіперпараметри, використані під час навчання

Гіперпараметр	Значення
Розмірність латентного простору для генератора	[100, 128, 256]
Коефіцієнт швидкості навчання	[$1 \cdot 10^{-3}$, $2 \cdot 10^{-4}$, $1 \cdot 10^{-4}$, $1 \cdot 10^{-5}$]
Розмір навчальної вибірки для ітерації навчання	[64, 128, 256]
Функція втрат	[<i>Binary Crossentropy, Wasserstein₁ distance</i>]
Функція активації	[<i>Linear, ReLU, LeakyReLU, TanH</i>]
Оптимізатор	[<i>SGD, Adam</i>]
Adam β_1 параметр	[0.0, 0.5]
Adam β_2 параметр	[0.9, 0.999]
Наявність повнозв'язних шарів	[<i>False, True</i>]
Вид нормалізації для згорткових шарів	[<i>None, Batch Normalization, Layer Normalization</i>]
Розмір ядра для згорткових шарів	[3, 4, 5]
Корекція градієнта	[<i>None, Gradient Clipping, Gradient Penalty</i>]
Ітерацій навчання генератора / дискримінатора за одну ітерацію навчання	[1 / 1 , 1 / 5]
Початковий розподіл ваг	[<i>GlorotNormal, Normal(0, 0.02), Uniform(-0.02 · $\sqrt{3}$, 0.02 · $\sqrt{3}$)</i>]
Використання зміщень	[<i>False, True</i>]

Після проведення навчання, вибору найкращих модифікацій та узагальнення було отримано такі результати (табл. 3).

ТАБЛИЦЯ 3. Порівняння результатів попереднього навчання модифікації GAN

	«DCGAN»	«DCGAN-WI»	«WGAN-GP»
Тривалість навчання	260 епох	60 епох	32 епох
Орієнтовний час навчання	280 хвилин	105 хвилин	530 хвилин
Час на одну епоху	65 секунд	105 секунд	995 секунд
Наявність «mode collapse»	Так	Так	Ні
Частота «mode collapse»	50 епох	30 епох	–
Фінальне значення FID	102.344	133.761	128.259

На основі отриманих результатів, можна зробити висновок, що оскільки основна мета аналізу різних модифікацій, була подолання проблеми «mode collapse» у процесі навчання, то як основну модель обрано модифікацію «WGAN-GP».

Значення функції втрат для мереж генератора та дискримінатора, що отримано в процесі навчання, показано на рис. 3.

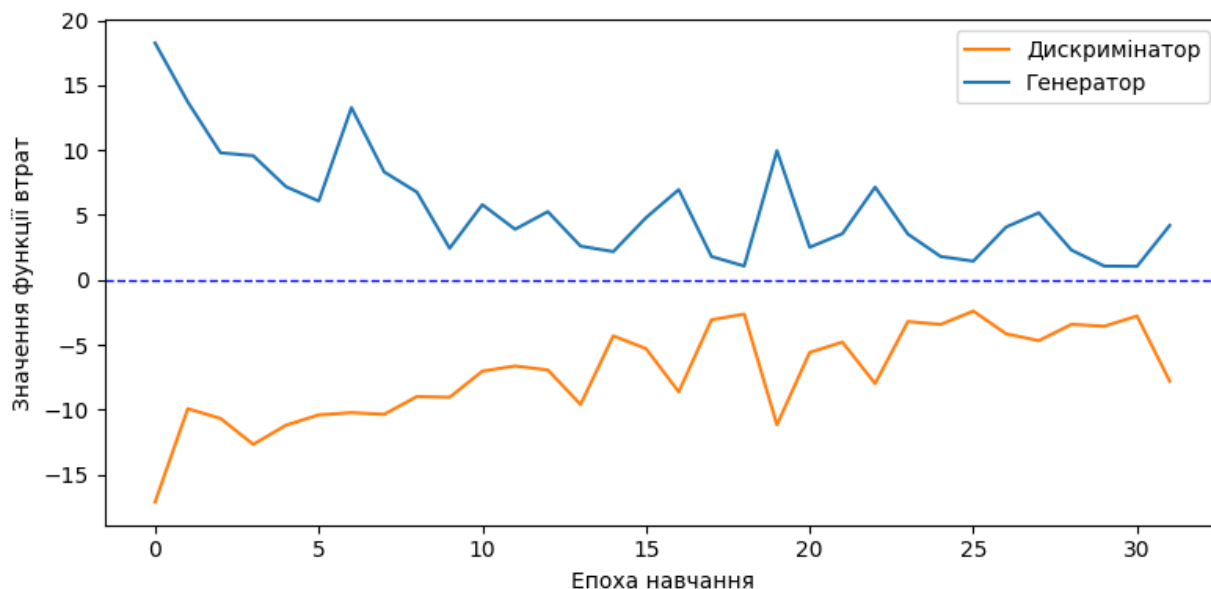


РИС. 3. Модель «WGAN-GP». Значення функції втрат для дискримінатора та генератора під час навчання

В свою чергу, значення оцінок дискримінатора для справжніх та згенерованих зображень, що отримано в процесі навчання, показано на рис. 4.

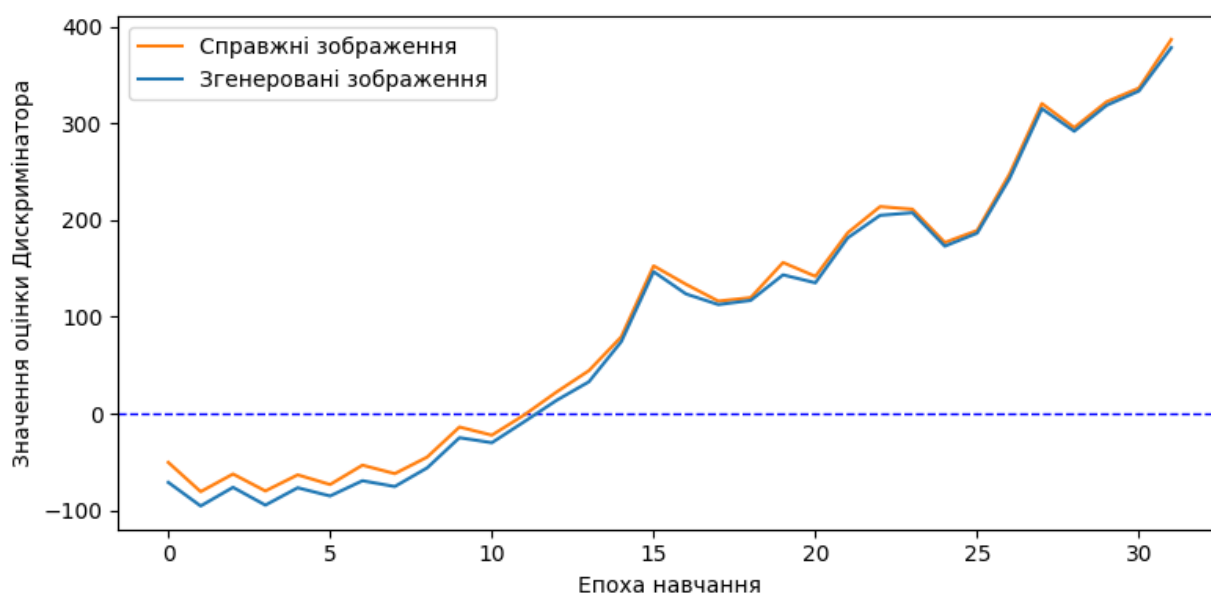


РИС. 4. Модель «WGAN-GP». Значення оцінки дискримінатора справжніх та згенерованих зображень під час навчання

Висновок. У роботі розроблено архітектуру та програмне забезпечення системи генерації зображень на основі методів глибокого навчання. Розроблена система може бути застосована для генерації зображень, що можуть бути використані як джерела натхнення для витворів мистецтва створених людиною тощо. Але, відповідна система має необхідний потенціал для повного виконання поставленої мети, повний процес створення витворів мистецтва. Для цього, система потребує подальшого вдосконалення та досліджень. Можливим напрямком модифікації системи є: тонший аналіз впливу гіперпараметрів на модель під час навчання; зміна структури моделей генератора та дискримінатора; підбір кращого навчального набору даних; аналіз початкових розподілів ваг у дискримінатора та генератора; зменшення коефіцієнту швидкості навчання та збільшення часу навчання.

Список літератури

1. Trach Y., 2021. Artificial Intelligence As A Tool For Creating And Analysing Works Of Art. *Culture and Arts in the Modern World*. 2021. 22. P. 164–173. <https://doi.org/10.31866/2410-1915.22.2021.235907>
2. Mehlig B. Machine learning with neural networks. Department of Physics University of Gothenburg Göteborg, Sweden. 2021. 241 p. <https://arxiv.org/abs/1901.05639> (звернення: 03.09.2023)
3. Goodfellow I. NIPS 2016 tutorial: Generative Adversarial Networks. 2017. <https://arxiv.org/abs/1701.00160> (звернення: 03.09.2023)
4. Brownlee J. A gentle introduction to generative adversarial networks (Gans). MachineLearningMastery.com. 2022. <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/> (звернення: 03.09.2023)
5. Radford A., Metz L., Chintala S. Unsupervised representation learning with deep convolutional generative Adversarial Networks. 2016. <https://arxiv.org/abs/1511.06434> (звернення: 03.09.2023)
6. Liao P., Li X., Liu X., Keutzer K. The ARTBENCH dataset: Benchmarking generative models with artworks. 2022. <https://arxiv.org/abs/2206.11404> (звернення: 03.09.2023)

Одержано 05.09.2023

Третиник Віолета Вікентіївна,

кандидат фізико-математичних наук,
доцент кафедри прикладної математики НТУУ «КПІ ім. Ігоря Сікорського»,
<https://orcid.org/0000-0002-3538-8207>
viola.tret@gmail.com

Будзінський Євген Олександрович,

магістрант кафедри прикладної математики НТУУ «КПІ ім. Ігоря Сікорського».

УДК 519.67

В.В. Третиник *, Є.О. Будзінський

Використання методів глибокого навчання для генерації зображень

НТУУ «КПІ ім. Ігоря Сікорського»

* Листування: viola.tret@gmail.com

Вступ. Художники завжди використовували різні засоби для вираження своєї творчості та дослідження своєї уяви. З розвитком цифрових технологій митці тепер мають доступ до величезної кількості інструментів, які дозволяють їм створювати витвори мистецтва, які є більш витонченими, складними та візуально захоплюючими, ніж будь-коли раніше. Останнім часом зростає інтерес до використання штучного інтелекту для створення зображень для художніх цілей.

Створення зображень для сфери мистецтва передбачає застосування алгоритмів і методів машинного навчання (МН) для створення цифрових творів мистецтва, які можуть імітувати стилі, техніки й естетику традиційних форм мистецтва. Системи штучного інтелекту можуть навчатися на величезних масивах даних, щоб створювати зображення, які є неймовірно реалістичними та деталізованими, а також унікальними та оригінальними.

Мета роботи. У даній роботі для генерації зображень було застосовано підхід, що базується на генеративних моделях. Методи машинного навчання, а саме: глибокі нейронні мережі відкривають широкі можливості для розв'язання поставленої задачі.

Результати. У даній роботі розглянуто застосування методів глибокого навчання для генерації зображень. Проведено порівняльний аналіз існуючих засобів генерації зображень. Запропонована модифікація генеративної моделі. Розроблена система генерує зображення фіксованого розміру (64x64, 256x256, 1024x1024) художнього характеру, мінімальне значення індексу FID, при навчанні – 128. Була виконана програма реалізація моделі на мові програмування Python.

Ключові слова: згорткові нейронні мережі, машинне навчання, генерація мистецтва, генеративна змагальна мережа.

UDC 519.67

Violeta Tretynuk *, Evgeny Budzinskyi

Using Deep Learning Methods for Image Generation

The National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

* Correspondence: viola.tret@gmail.com

Introduction. Artists have always used different mediums to express their creativity and explore their imagination. With the advancement of digital technology, artists now have access to a vast array of tools that allow them to create works of art that are more sophisticated, complex, and visually engaging than ever before. Recently, there has been a growing interest in using artificial intelligence to create images for artistic purposes.

Art imaging involves the application of algorithms and machine learning (ML) techniques to create digital works of art that can mimic the styles, techniques, and aesthetics of traditional art forms. Artificial intelligence systems can learn from massive amounts of data to create images that are incredibly realistic and detailed, as well as unique and original. The purpose of the paper. In this paper, an approach based on fuzzy logic was used

to estimate the cost of housing in Kyiv. Fuzzy methods allow to apply a linguistic description of complex processes, to establish fuzzy relationships between concepts, to predict the behavior of the system, to create a set of alternative actions, to formally describe fuzzy decision-making rules.

The purpose of the article. In this work, an approach based on generative models was used for image generation. Machine learning methods, namely deep neural networks, open wide opportunities for solving the given problem.

Results. This paper considers the application of deep learning methods for image generation. A comparative analysis of existing means of image generation was carried out. A proposed modification of the generative model. The developed system generates an image of a fixed size (64x64, 256x256, 1024x1024) of an artistic nature, the minimum value of the FID index during training is 128. The program implementation of the model was performed in the Python programming language.

Keywords: convolutional neural networks, machine learning, art generation, generative competitive network.