УДК 004:005:51

СУЧАСНІ ТЕХНОЛОГІЇ АЛГОРИТМІЗАЦІЇ І МОДЕЛІ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ

А. А. Ковальчук, Н. А. Потапова

Анотація. У статті проаналізовано сучасні технології генерації зображень за допомогою штучного інтелекту. У роботі пояснюються принципи роботи моделей DALL·E, Midjourney і Stable Diffusion, зокрема їх алгоритми (трансформери, дифузійні моделі) і структури даних (тензори, векторні представлення). Подано коротку історію розвитку текст-ту-імейдж генерації та її сучасний стан, від GAN-методів до дифузійних мереж. Виконано порівняльний аналіз трьох інструментів за критеріями: якість зображень, швидкість генерації, доступність, гнучкість налаштувань та ліцензійні умови (open-source). Наведено переваги й недоліки кожного підходу на основі реальних джерел і оглядів.

Ключові слова: генерація зображень, штучний інтелект, генеративні моделі, трансформер, дифузійна модель.

Вступ. Зі зростанням обсягів даних та потужності обчислювальних систем, автоматичне створення реалістичних зображень із текстового опису стало важливим напрямом штучного інтелекту. Формулювання «текст-до-зображення» (text-to-image) передбачає, що система генерує візуальний контент за словесною інструкцією. Як зазначають дослідники, «проєктування автоматичної системи, яка генерує реалістичні зображення з описів мовою, є нетривіальним завданням і значним кроком на шляху до штучного інтелекту, подібного до людського» [1]. Початкові методи ґрунтувалися на генеративних змагальних мережах (GAN) та варіаційних автоенкодерах (VAE), але вони мали обмеження у якості та керованості. Важливим етапом стали авторегресивні трансформери, такі як DALL·E від OpenAI, що демонстрували можливість безпосередньо генерувати зображення на основі мови. Однак вони вимагали величезних обчислювальних ресурсів. Нещодавно з'явився новий «суспільний стандарт» – дифузійні моделі, які досягли найкращих результатів у генерації зображень [2].

Метою статті є аналіз основних підходів, алгоритмів та структур даних, що лежать в основі технологій генерації зображень, та порівняння сучасних систем: DALL·E, Midjourney та Stable Diffusion.

Основна частина. Модель DALL·E (OpenAI). DALL·E (названий від імені художника Сальвадора Далі та персонажа WALL-E) – перша велика система OpenAI для генерації зображень з текстових описів. За офіційними даними, це трансформерна модель з 12 мільярдами параметрів, побудована на архітектурі GPT-3 і навчена на великому наборі пар «текст–зображення» [3], [4]. Під час навчання DALL·E отримує текстове повідомлення та зображення як єдиний послідовний потік токенів, після чого прогнозує всі токени послідовності, включаючи частини зображення [5]. Такий підхід дозволяє моделі не лише створювати нове зображення «з нуля» за текстом, але й дорисовувати частину існуючого зображення, узгоджуючись із запитом. Водночас авторегресивна природа трансформера означає, що кожний новий піксель (токен) генерується послідовно, що робить процес витратним за часом обчислень.

Незважаючи на інноваційність, DALL·E першої версії мала обмеження у якості та роздільній здатності. У наступних версіях (DALL·E 2 і 3) OpenAI перейшов на дифузійні підходи з використанням CLIP (Contrastive Language–Image Pretraining). Наприклад, DALL·E 2 використовує модифіковану GLIDE-дифузію: текстовий запит спочатку перетворюється у векторне уявлення зображення за допомогою CLIP, а потім дифузійний декодер поступово відновлює повноцінне зображення. Отже, головні алгоритми DALL·E – це трансформери (для генерації і попереднього завдання) і дифузійні декодери (для підвищення якості образів) [8].

Модель Stable Diffusion. Stable Diffusion – відкрита система для генерації зображень на основі дифузійних моделей, розроблена командою CompVis за підтримки Stability AI. Це латентна дифузійна модель (LDM): перед генеруванням зображення він спочатку стискає його в нижчий латентний простір за допомогою варіаційного автоенкодера (VAE), а потім застосовує процес дифузії до цього стислого представлення. У Stable Diffusion архітектура складається з трьох основних компонентів: VAE-енкодера / декодера, UNet для послідовного відновлення

та опціонального текстового енкодера (зазвичай CLIP) [8]. Під час генерації спочатку формується випадковий шум у латентному просторі, а потім UNet з ResNet-блоками поступово «дешумізує» його протягом низки кроків, керуючись вектором текстового запиту (через механізм крос-атенції з CLIP) [10].

У працях CompVis було показано, що тренування дифузійних моделей у латентному просторі значно знижує обчислювальну складність, зберігаючи високу якість зображень. Запровадження шарів крос-атенції у моделі дозволило ефективно враховувати текстове керування під час генерації [7]. На відміну від DALL·E, Stable Diffusion має відкритий код і ваги моделі, що можна запускати навіть на звичайних GPU з невеликим об'ємом пам'яті [8].

Модель Midjourney. Midjourney – комерційний сервіс генерації зображень, доступний переважно через Discord та вебінтерфейс. Хоча точна архітектура Midjourney не розкривається публічно, вважається, що це видозмінена дифузійна модель, імовірно, похідна від Stable Diffusion з додатковим файнт'юнінгом і власною системою стилізації. У роботі Ticong (eWeek) зазначено, що Midjourney спеціалізується на деталізованих і стильних образах із потужними налаштуваннями (зміна роздільності, стилізації, upscale тощо) [15]. Тобто основними алгоритмами під капотом є дифузія (ітеративне видалення шуму) та навряд чи відмінні від Stable Diffusion механізми самоконтролю, хоча точні параметри залишаються закритими.

У всіх перелічених моделях основу складають глибинні нейронні мережі, які є графами зв'язаних шарів із матричними операціями. Дані (текстові підказки, образи) представляються у вигляді тензорів (багатовимірних масивів чисел). Наприклад, текст переводиться у вектори-ембедінги (слово = числовий вектор), зображення представлено як тривимірний тензор пікселів, а внутрішні стани мережі – також як тензори. Алгоритмічно ключовими є трансформерні механізми з self-attention (у DALL·E) та алгоритми дифузії (марківська послідовність зворотного зняття шуму у StableDiffusion / Midjourney) [2].

Розпізнавання тексту та зв'язок із візуальним вмістом забезпечує модель CLIP, що переводить текст у «психометричний» простір, де є відповідність із зображеннями. З точки зору структур даних, генерація зображень використовує вектори з декількох сотень чи тисяч розмірів для представлення семантики. Алгоритми навчання – це, зокрема, стохастичний градієнтний спуск (оптимізація параметрів мережі) із метою мінімізації невизначеності (лог-лікелігуд), зважаючи на шум. Під час генерації Stable Diffusion і Midjourney проходять близько сотні кроків дифузії, тобто алгоритмічно здійснюється марківський процес, який поступово очищує шум (зворотне дифузійне моделювання) [6].

Насамкінець коротко зазначимо, що розвитку цих технологій передували GAN-моделі (наприклад, conditional GAN для зображень із текстом), але їм поступово на зміну прийшли трансформери та дифузійні схеми, що довели свою ефективність [2]. Зараз на піку розвитку знаходяться версії DALL·E 3, Midjourney v6/7 та Stable Diffusion XL (2023–2024), які генерують образи дуже високої якості, і інтегруються у різні платформи (чат-боти, мобільні додатки тощо).

Проведемо порівняльний аналіз інструментів:

– Якість зображень. Midjourney часто генерує найбільш деталізовані та художні результати серед інших, проте вони можуть містити неточності щодо конкретних деталей запиту. У експериментах Stable Diffusion показує більш високу відповідність текстовому опису: наприклад, якщо запит містить конкретні елементи (скляні конструкції, внутрішнє наповнення), Stable Diffusion зазвичай їх чітко відтворює, тоді як Midjourney іноді «пропускає» такі деталі заради загальної естетики. DALL·E (з останніми версіями) дає добре збалансовані зображення: вони й детальні і близькі до опису, але зазвичай їх творча виразність трохи поступається Midjourney. Загалом за художньою виразністю Midjourney, за точністю та стабільністю результату – Stable Diffusion, а за оптимальним балансом – DALL·E.

– Швидкість генерації. Усі моделі дозволяють отримати картинку у відносно короткий час (від кількох секунд до хвилини) залежно від потужності обладнання та обраної роздільної здатності. DALL·E 3 на платформі ChatGPT генерує зображення практично миттєво (в межах кількох секунд), завдяки хмарним серверам OpenAI. Midjourney також демонструє швидке створення, проте залежить від черги в Discord: зазвичай зображення з'являється за 10–30 се-

кунд. Stable Diffusion може працювати швидко на сучасному GPU, але у разі локального запуску на слабшому залозі або CPU-версії процес може затягуватися (декілька десятків секунд на картинку). Важливо, що Stable Diffusion дозволяє обирати різну кількість ітерацій дифузії (кроків), що прямо впливає на швидкість: менше кроків – швидше результат, але потенційно нижча якість.

– Доступність та гнучкість. Stable Diffusion є найгнучкішим у доступі: її відкритий код і моделі доступні безліччю платформ. Наприклад, генерацію можна запускати локально на власному комп'ютері, а також через вебсервіси DreamStudio (від Stability AI), Hugging Face Spaces або чат-бот Stable Assistant. Завдяки цьому Stable Diffusion не вимагає постійного інтернет-з'єднання та дозволяє тонке налаштування (fine-tuning) моделі. У DALL·E 3 доступ реалізовано через декілька каналів: основний – через інтеграцію з ChatGPT на веб та мобільних додатках, а також через платний API OpenAI [5, 6, 8]. Це забезпечує зручний інтерфейс (чат) і вибір платформ, але самі моделі не можна запустити локально. Midjourney доступний лише через Discord або власний вебінтерфейс – це обмежує варіанти доступу (потрібне підключення до інтернету та обліковий запис Discord). Отже, Stable Diffusion – найдемократичніша у доступі, DALL·E – помірно доступна (через чат-бот), а Midjourney – найбільш обмежена.

– Гнучкість налаштувань. Midjourney і Stable Diffusion дозволяють змінювати стиль та якість зображення за допомогою параметрів. У Midjourney є вбудовані інструменти для масштабування, зуму, перегенерації за тим же описом тощо [5]. До того ж у Midjourney можна вказувати стилі (через текстові параметри) і навіть корегувати зображення після генерації (вбудований редактор). Stable Diffusion надає ще ширші можливості: її можна до-навчати (fine-tune) на своїх даних, використовувати різні моделі (SDXL, SD3.0 тощо), стилістичні гіпермережі (hypernetworks) для перенесення стилю, а також обирати різні алгоритми семплінгу (PLMS, ddim, Euler та ін.) для балансу швидкість–якість [21]. Багато платформ з SD (DreamStudio, AUTOMATIC1111, InvokeAI тощо) мають численні налаштування: вибір семплеру, CFG scale, ширина контексту і т. д. DALL·E 3 найменш гнучкий щодо налаштувань: він пропонує простий інтерфейс (переважно опис у чаті), але не дає користувачу прямого доступу до параметрів генерації – все відбувається у бекенді OpenAI. Отже, Stable Diffusion виявляє максимальну гнучкість, Midjourney – високу, а DALL·E – обмежену з точки зору ручного налаштування.

– Ліцензії та відкритий код. Stable Diffusion створена на основі відкритих моделей: її вихідні коди та ваги можна завантажити і використовувати безкоштовно для неприбуткового та малобізнесового використання. За умовами Stability AI, безкоштовно можна використовувати Stable Diffusion у власних проєктах, якщо річний дохід організації менший за 1 млн дол. США. Також користувачі зберігають повні права на створені ними зображення. Натомість DALL E та Midjourney – закриті пропрієтарні сервіси: їх моделі не можна запустити самостійно, потрібен доступ до хмарних сервісів компаній. У DALL·E (OpenAI) генерація зображень відбувається за ліцензійними умовами OpenAI, хоча OpenAI заявляє, що користувач отримує права на створені ним зображення. Midjourney працює за підпискою (стандарт, преміум тощо) і також зберігає авторські права на контент певною мірою (наприклад, базовий рівень дозволяє іншим користувачам бачити згенеровані вами картинки). Як зазначено у джерелах, перевагою Stable Diffusion є відверті ліцензійні умови (Community License) та відкритість, на противагу більш жорсткому підходу Midjourney і DALL·E.

Висновки. Генерація зображень за допомогою штучного інтелекту нині відбувається на межі технологічних можливостей. Теоретично ці системи базуються на складних алгоритмах – глибоких нейронних мережах, трансформерах та дифузійних моделях – та оперують величезними масивами даних (тензорами пікселів і ембедінгів). Історично спостерігалося поступове зростання складності: від перших GAN до авторегресивних трансформерів (DALL E) і тепер – до дифузійних моделей (Stable Diffusion, Midjourney). Практично кожна з розглянутих систем має свої плюси. Midjourney виграє за художньою виразністю і багатством стилів, Stable Diffusion – за доступністю та відкритим кодом, а DALL·E 3 – за простотою інтеграції у чат-інтерфейси і гарантіями якості від OpenAI. Кожен інструмент підходить для різних сце-

наріїв: для ідеї та натхнення (Midjourney), для експериментів і кастомізації (Stable Diffusion) або для швидкої та надійної генерації через хмару (DALL·E). З огляду на швидкий розвиток галузі, майбутні моделі, ймовірно, ще більше покращать якість і швидкість, а також розширять можливості керування процесом генерації.

Annotation. This article analyzes modern AI-based image generation technologies. We explain how models like DALL·E, Midjourney, and Stable Diffusion work (transformers, diffusion, neural architectures), including data structures (tensors, embeddings). A brief history from GANs to current diffusion models is given. In the practical part, we compare the three tools in terms of image quality, speed, accessibility, flexibility, licensing and open-source status. The strengths and weaknesses of each approach are highlighted. Keywords: image generation, artificial intelligence, generative models, DALL·E, Midjourney, Stable Diffusion, transformer, diffusion model, CLIP, open source.

Keywords: image generation, artificial intelligence, generative models, transformer, diffusion model.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. DALL·E: Creating images from text. *OpenAI.* 2021. URL: https://openai.com/index/dall-e/
2. Text-to-image Diffusion Models in Generative AI: A Survey / Chenshuang Zhang, Chaoning Zhang, M. Zhang, In So Kweon. 2024. URL: https://ar5iv.labs.arxiv.org/html/2303.07909
3. Rombach R., Blattmann A., Lorenz D. High-Resolution Image Synthesis with Latent Diffusion Models. *CVPR.* 2022. P. 10684–10695. URL: https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html
4. Stable Diffusion. *Wikipedia The Free Encyclopedia.* URL: https://en.wikipedia.org/wiki/Stable_Diffusion
5. Ticong L. Midjourney vs DALL-E: AI Art Tools Face-Off for 2025. *eWeek.* 2025. URL: https://www.eweek.com/artificial-intelligence/midjourney-vs-dalle/
6. Ticong L. Midjourney vs Stable Diffusion: 2025's Creative Clash. *eWeek.* 2025. URL: https://www.eweek.com/artificial-intelligence/midjourney-vs-stable-diffusion/
7. Self-Hosted License. *Stability AI.* URL: https://stability.ai/license
8. Ryan O'Connor. How DALL-E 2 Actually Works. *Asamblyai.* 2023. URL: https://www.assemblyai.com/blog/how-dall-e-2-actually-works

УДК 004:005:51

АЛГОРИТМИ НА ГРАФАХ ТА ЇХ ЗАСТОСУВАННЯ

Д. О. Круцюк, Н. Р. Веселовська

Анотація. У публікації розглянуто ключові алгоритми на графах та їх фундаментальну роль у розв'язанні сучасних обчислювальних задач. Проаналізовано базові підходи до обробки графових структур, зокрема алгоритми пошуку в ширину (BFS) та глибину (DFS), а також методи знаходження найкоротших шляхів, зокрема алгоритм Дейкстри. Особливу увагу приділено практичному застосуванню цих алгоритмів у реальних системах: від GPS-навігації та оптимізації логістики до аналізу соціальних мереж.

Ключові слова: граф, алгоритм, оптимізація, пошук у ширину, Дейкстра, маршрутизація, аналіз мереж.

Вступ. У сучасному світі, що побудований на зв'язках – від соціальних мереж до глобальних логістичних маршрутів, алгоритми на графах стали ключовим інструментом для аналізу та оптимізації. Вони дозволяють розв'язувати складні задачі, де головну роль відіграє не сама сутність об'єктів, а структура відносин між ними. Саме ці алгоритми є в основі GPS-навігаторів, систем рекомендацій та інструментів для виявлення загроз у комп'ютерних мережах, які роблять наше життя більш ефективним.

Хоча теорія графів має глибоке математичне коріння, її практичне значення значно зросло з розвитком цифрових технологій. Сьогодні ці методи знаходять застосування у найрізноманітніших галузях: від біоінформатики, де графи моделюють взаємодії білків, до аналізу фінансових ринків. За допомогою графових алгоритмів можна не тільки обробляти величезні мережеві структури, але й знаходити в них приховані закономірності та найоптимальніші рішення.

У цій публікації ми розглянемо фундаментальні алгоритми на графах, їх основні принципи роботи та приклади застосування. Особливу увагу буде приділено тому, як методи обходу та пошуку найкоротшого шляху впливають на розвиток сучасних інформаційних систем.