Розробка методу генерації зображень на основі заданого текстового опису для обходу водяних знаків

УДК 621.395.7 (043.2)

Максим Житніков[1], Дмитро Пелешко[1,2], Олена Винокурова[2]

[1]Національний університет "Львівська Політехніка", maksym.zhytnikov.mknssh.2023@lpnu.ua
[2]Львівський національний університет ім. Івана Франка, dmytro.peleshko@lnu.edu.ua, olena.vynokurova@lnu.edu.ua

Водяні знаки - це поширений метод захисту авторських прав і прав власності на цифрові зображення шляхом вбудовування ледь помітного знаку або візерунка, який ідентифікує автора або законного власника. Однак цей метод іноді може бути перешкодою, особливо у випадках, де візуальна цілісність зображення має першорядне значення. Тому зростає інтерес до розробки технологій, які можуть генерувати зображення, обходячи водяні знаки або захищаючи від них [1]. Генеративні алгоритми можуть створювати абсолютно нові зображення. Ця здатність робить їх ідеальними для створення зображень, які природним чином уникають будь-яких водяних знаків, які зазвичай наносяться на оригінальні роботи.

Метод генерації зображень на основі заданого текстового опису може зробити значний внесок у вирішення проблеми обходу цифрових водяних знаків кількома способами. По-перше, ці архітектури призначені для генерування нових зображень з текстових описів [2]. Цей процес за своєю суттю дозволяє уникнути прямого дублювання будь-яких існуючих зображень, на які можуть бути нанесені водяні знаки. Оскільки згенеровані зображення є оригінальними творіннями, вони не переносять жодних вбудованих водяних знаків, по суті, обходячи водяний знак без необхідності його видалення. По-друге, гнучкість і творчий підхід, які пропонують моделі перетворення тексту в зображення, дозволяють створювати зображення таким чином, що водяні знаки стають неефективними. Наприклад, якщо водяний знак зазвичай розміщується в певній області зображення, модель можна навчити генерувати зображення з ключовими візуальними елементами, що займають ці області, природно, приховуючи будь-який водяний знак, який може бути доданий пізніше. Загалом, технологія генерації зображень на основі заданого тексту пропонує універсальне рішення проблеми цифрового водяного маркування шляхом створення оригінального контенту, який не порушує права авторів оригінальних зображень, зберігаючи при цьому естетичні та функціональні якості зображень.

Доволі популярним вирішенням проблеми генерації зображень на основі заданого тексту є архітектура [3]. Запропонований метод підтримує генерацію зображень на основі заданого класу (conditional generation) та сегментації-маски. Основним її недоліком є відсутність можливості генерації зображень на основі заданого вхідного тексту.

У даній роботі пропонується модифікація архітектури [3] таким чином, що стає можливою генерація зображень на основі заданого тексту. На рисунку 1, представлена модифікована архітектура, яка дозволить генерувати зображення на основі заданого тексту.

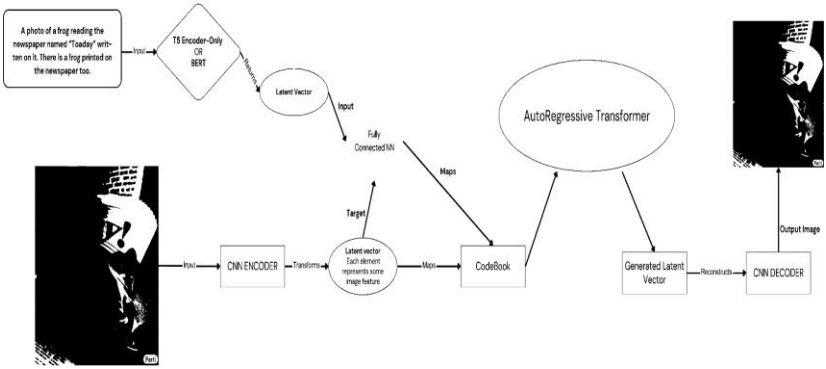

Рис.1. Модифікація архітектури [1]

На рисунку 1 відображається модульна архітектура, яка поєднює авторегресивний трансформер та VQ-GAN [3, 4]. VQ-GAN складається з двох частин: VQ-VAE в ролі генератора, та класичного конволюційного дискримінатора. Генерація зображення відбувається шляхом трансформації вхідного зображення, використовуючи конволюційний енкодер, у латентне векторне представлення. Кожен елемент латентного вектора відображає певну характеристку-особливість зображення, таку як, наприклад, наявність ока чи голови. Далі, цей латентний вектор співставляється із codebook [3, 5], яка була отримана в результаті тренування VQ-GAN. Дана codebook містить дискретні латентні вектори, кожен з яких відповідає за певний елемент який може бути представлений на зображенні.

Отримані латентні вектори співставляються з найближчими дискретними латентними векторами з codebook. Далі на основі цих латентних дискретних кодів, трансформер авторегресивно генерує зображення, представлене у вигляді латентних кодів-харектеристик, які подані як дискретні вектори у codebook [2]. Цей авторегресивний підхід, де генерація кожної частини зображення послідовно залежить від попередньо згенерованих частин, забезпечує послідовність і контекстуальну узгодженість вихідних даних. Далі конволюційний декодер генерує вже повноцінне зображення, використовуючи латентні вектори характеристик зображення, які згенеровані авторегресивним трансформером.

Під час генерації зображення на основі вхідного текстового опису, використовується енкодер-частина T5 text-to-text трансформера, який перетворює вхідний текстовий опис у латентний вектор, який, у свою чергу, чітко охоплює всі текстові залежності та деталі наявні у вхідному текстовому описі. Далі використовується натренована повнозв'язна нейронна мережа, яка трансформує латентний вектор вхідного текстового опису, отриманий від T5 енкодера, у формат, співставний із латентним вектором, отриманим від конволюційного енкодера. Таким чином, латентний вектор зображення і латентний вектор відповідного

текстового опису будуть приблизно однаковими. Після цього латентний вектор текстового опису співставляється із дискретними латентними векторами із codebook, подається у трансформер, який, у свою чергу, авторегресивно генерує дискретні латентні вектори вихідного зображення. В результаті, конволюційний декодер реконструює повноцінне вихідне зображення.

Розроблений метод генерації зображень на основі текстового опису показує великий потенціал у обході цифрових водяних знаків. Використання комбінації генеративних алгоритмів, таких як T5, VQ-GAN та авторегресивного трансформера, дозволяє створювати візуально привабливі та унікальні зображення, що ефективно уникають стандартних методів водяного маркування. Це становить значний крок у забезпеченні цифрової безпеки зображень, одночасно зберігаючи їх естетичну цінність і доступність для широкої публіки.

1. Huiwen Chang, Han Zhang, Jarred Barber, AJ Maschinot, Jose Lezama, *et al.* Muse: Text-To-Image Generation via Masked Generative Transformers arXiv, Jan. 2, 2023. Accessed: Apr. 29, 2024. [Online]. Available: https://arxiv.org/abs/2301.00704
2. Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, *et al.*, Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. arXiv, Jun. 21, 2022. Accessed: Apr. 21, 2024. [Online]. Available: http://arxiv.org/abs/2206.10789
3. Patrick Esser, Robin Rombach, Björn Ommer Transformers for High-Resolution Image Synthesis. arXiv, Jun. 23, 2021. Accessed: Apr. 21, 2024. [Online]. Available: http://arxiv.org/abs/2012.09841
4. Shiyue Cao, Yueqin Yin, Lianghua Huang, Yu Liu, Xin Zhao, Deli Zhao and Kaiqi Huang., Efficient-VQGAN: Towards High-Resolution Image Generation with Efficient Vision Transformers. arXiv, Oct. 09, 2023. Accessed: Apr. 21, 2024. [Online]. Available: http://arxiv.org/abs/2310.05400
5. Long Zhao, Zizhao Zhang, Ting Chen, Dimitris N. Metaxas, Han Zhang. Improved Transformer for High-Resolution GANs. arXiv, Dec. 23, 2021. Accessed: Apr. 21, 2024. [Online]. Available: http://arxiv.org/abs/2106.07631