

УДК 004.8

## МОДИФІКАЦІЯ STYLEGAN ДЛЯ СТВОРЕННЯ РЕАЛІСТИЧНИХ ВИСОКОТОЧНИХ ЗОБРАЖЕНЬ

Чухран І. Д.

Науковий керівник – ст. викл. Гриньова О. Є.

Харківський національний університет радіоелектроніки, каф. ШІ  
м. Харків, Україна

тел.: +38(099) 382 07 61, e-mail: ivan.chukhran@nure.ua

Comparative analysis Generative Adversarial Networks, the StyleGAN architecture are reviewed. The modification of StyleGAN for distinguish high-resolution image generation using a style-based generator architecture with unsupervised style mapping is proposed. The system testing is performed.

StyleGAN – різновид Generative Adversarial Network (GAN), особливістю якого є стилі – специфічні ознаки зображуваного об'єкта (наприклад, поза та вираз обличчя людини на фото). StyleGAN та її варіації стали революційними для створення реалістичних високоточних зображень, хоча й мали деякі недоліки, які були вирішені більш молодими версіями типу StyleGAN2.

Загальна структура такої моделі схожа на базову й складається з генератора (G) – нейронна мережа, яка генерує шумове зображення, яке після навчання стає схожим на реальне; дискримінатора (D) – нейронна мережа, яка визначає, чи є зображення, яке надходить від генератора, реальним, чи воно штучно згенеровано; критерію (Cr) – функція, яка визначає, на скільки відрізняються значення  $x_1$  та  $x_2$ ; та оптимізатора (Opt) – механізм, який змінює параметри генератора та дискримінатора відповідно до значень, отриманих з критерію.

Замість простої генерації вектору випадкових значень, згідно нормального розподілу  $z \in Z$ , й їх передачу до G було вирішено створити проміжний стан  $w \in W$  шляхом нелінійної обробки через Mapping Network  $f: Z \rightarrow W$ . Це 8-шарова модель без вчителя, яка розміщує стилі у латентному просторі. Це дозволяє виокремити конкретні стилі та запобігти сильній кореляції між ними. Такий механізм не тільки звільняє від вбудовування окремого вектору стилів у вхідний шумовий вектор, що значно спрощує подальшу обробку даних [1].

Далі  $w$  використовується для змішування стилів разом з новими шумовими векторами, які додаються після кожної операції в блоці G. Кожний такий блок складається з таких елементів: Upscaler – функція інтерполяції розмірів для забезпечення високої якості вихідного зображення; транспонована згортка та шар нормалізації Adaptive Instance Normalization (AdaIN), яка надає можливість коригувати рівень накладання стилю на кожний з каналів зображення.

Задля підтримки стохастичної варіативності результуючих зображень було вирішено по-піксельно додавати випадкові значення згідно закону

нормального розподілу – це гарантує унікальність деталей зображення. Наявність AdaIN в кожному блоці також підтримує стохастичність після кожного додавання шумових значень до кожного стилю. На відміну від інших GAN, запропонована мережа використовує критерій Frechet Inception Distance (FID), FID використовують для знаходження відстані між розподілами двох чи більше випадкових величин. Вибір зумовлений необхідністю розрахунку відстані між реальними та згенерованими зображеннями згідно їхнього розподілу, на відміну від звичної Cross-Entropy Loss, яку використовували Deep Convolutional GANs.

Однією з головних особливостей представленої моделі є змішування стилів за допомогою вектору стилів  $w$ , завдяки якому можна переносити ключові ознаки одного зображення на інші шляхом інтерполяції між ними. Окрім використання лінійної інтерполяції було надано використання сферичної інтерполяції, оскільки створений латентний простір може бути викривленим в проміжних точках лінійної інтерполяції.

Обраний підхід не дає змоги контролювано створювати зображення за поданими критеріями, як це можливо у BigGAN (великомасштабна GAN для високоточного синтезу природних зображень), але дозволяє надавати необхідні варіанти зображень шляхом змішування стилів та їхнього відсіювання. Використання методів навчання зі вчителем дозволяють краще розмежувати стилістичні особливості зображення, аніж власноручне створення вкладених векторів особливостей, яке витрачає багато ресурсів.

Хоча мережа й здатна генерувати зображення високої якості, вона містить ряд недоліків. Найпоширенішим із них є наявність артефактів [2] на створеному зображенні через накладання інформації пікселів під час розгортки та нормалізації вихідного зображення. Проблема вирішується ревізією архітектури – якщо винести Upscaler за межі кожного стильового блока  $G$  та провести модуляцію конволюції (переобчислення вагових коефіцієнтів згортки згідно нормалізованих значень отриманих векторів), то можна позбутись наявних шумових викривлень. Додатково можна замінити лінійну інтерполяцію на інтерполяцію найближчих сусідів, що може зменшити вплив сусідніх пікселів при деконволюції.

Результати дослідження – проектування модифікованої StyleGAN та програмна реалізація демоверсії, тестування за критеріями: швидкість генерації, чіткість зображення, наявність артефактів, відповідність та різноманітність згенерованої вибірки.

Список використаних джерел:

1. Karras, T., Laine, S., & Aila, T. (2020). A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, <https://doi.org/10.1109/tpami.2020.2970919>.

2. Karras, T., Laine, S., Aittala, M., Hellsten, J., (2020). Analyzing and improving the image quality of stylegan. 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE. <https://doi.org/10.1109/cvpr42600.2020.00813>.