

РОМАН ПЕЛЕЩАК

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-0536-3252>e-mail: roman.m.peleshchak@lpnu.ua**КОПАЧ БОГДАН**

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0002-5158-589X>e-mail: bohdan.v.kopach@lpnu.ua**ІВАН ПЕЛЕЩАК**

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-7481-8628>e-mail: ivan.r.peleshchak@lpnu.ua

ГЕНЕРАЦІЯ СУПУТНИКОВИХ ЗОБРАЖЕНЬ НА ОСНОВІ ТЕКСТУ ЗА ДОПОМОГОЮ НЕЙРОННОЇ МЕРЕЖІ GAN ПОЄДНАНОЇ З ТРАНСФОРМЕРОМ SBERT

Генерація супутникових зображень є складним завданням, яке передбачає не лише опрацювання великих обсягів даних, а й потребує аналізу додаткових факторів, які впливають на якість та деталізацію зображення, зокрема географічної різноманітності, діяльності людей, змін пір року, погоди тощо. Використання GAN мереж дозволяє враховувати ці аспекти та моделювати багаторівневі складні супутникові фотознімки. Актуальність дослідження полягає в аналізі того, як використання GAN мереж у поєднанні із сучасними мовними моделями впливає на різноманітність, деталізацію та відповідність згенерованих супутникових зображень до реальних даних. Це особливо важливо у завданнях моніторингу, геопросторового аналізу, міського планування, адже допомагає швидко отримувати наближені до реальних високоякісні зображення відсутніх чи складних для збору ділянок територій та розширювати тренувальні вибірки для інших моделей.

Головна мета цієї статті – розробка та експериментальна перевірка точності GAN моделей із кодувальником тексту на основі SBERT, оцінка керованості моделей та перевірка того, наскільки використання цього підходу дозволяє правильно опрацьовувати семантичну інформацію із текстових описів та перетворювати її в реалістичні зображення. Отримані результати показують, що модифікації згорткових GAN мереж (DCGAN та WGAN-GP) у поєднанні із SBERT здатні із великою точністю перетворювати текстові дані в супутникові зображення. Визначено, що збільшення кількості епох під час навчання є критично важливим, адже супутникові зображення вирізняються значною складністю та різноманітністю, а текстові описи – варіативністю стилів і контекстів. Запропонований метод інтеграції кодувальника на основі SBERT є універсальним рішенням, яке можна використовувати для навчання різних типів GAN моделей.

Ключові слова: нейронні мережі, GAN, SBERT, генерація зображень, супутникові зображення.

PELESHCHAK ROMAN, BOHDAN KOPACH, PELESHCHAK IVAN

Lviv Polytechnic National University

GENERATION OF SATELLITE IMAGES FROM TEXT USING GAN NEURAL NETWORK COMBINED WITH AN SBERT TRANSFORMER

Satellite image generation is a complex task that involves not only processing large volumes of data but also analyzing additional factors that affect image quality and detail, such as geographic diversity, human activity, seasonal changes, and weather conditions. The use of GAN networks enables consideration of these factors and allows for modeling multi-layered, complex satellite imagery. The relevance of this study lies in examining how the application of GAN networks in combination with modern language models impacts the diversity, level of detail, and fidelity of generated satellite images relative to real-world data. This is especially important for tasks in monitoring, geospatial analysis, and urban planning, as it facilitates the rapid production of high-quality images that closely approximate real data for areas that are missing or difficult to collect, and also helps expand training datasets for other models.

The main objective of this paper is to develop and experimentally verify the accuracy of GAN models with an SBERT-based text encoder, to assess their controllability, and to evaluate how effectively this approach processes semantic information from textual descriptions and transforms it into realistic images. The results demonstrate that modifications of convolutional GAN networks (DCGAN and WGAN-GP), when combined with SBERT, can convert textual data into satellite imagery with high accuracy. It was also determined that increasing the number of training epochs is critically important, given the substantial complexity and diversity of satellite images and the variability of text descriptions in terms of style and context. The proposed method of integrating the SBERT-based encoder is a universal solution that can be applied to train various types of GAN models.

Keywords: neural networks, GAN, SBERT, image generation, satellite images.

Постановка проблеми

В останні роки генеративні нейронні мережі різних типів активно використовуються для розв'язання різноманітних завдань, зокрема створення музичних композицій, зображень, текстів. Поява GAN моделей у 2014 році [1] стала важливим етапом розвитку генеративних мереж, оскільки використання двох нейронних мереж, які тренуються паралельно (генератора та дискримінатора), дозволило значно покращити ефективність навчання, якість згенерованих даних і стало стимулом для нових розробок та досліджень у царині штучного інтелекту.

Однією із сфер, де нейронні мережі набули значного поширення, є дистанційне зондування Землі та робота із супутниковими знімками. Генерування супутникових фотографій за допомогою моделей на кшталт GAN дає змогу суттєво розширити можливості обробки даних про нашу планету та створює нові перспективи для багатьох галузей. Нейронні мережі можуть використовуватися для відновлення або покращення супутникових зображень, створення карт, виявлення або усунення необхідних об'єктів тощо

[2]. Чимало завдань у сфері дистанційного зондування вимагають великих наборів різноманітних та збалансованих тренувальних даних (наприклад, для навчання моделей класифікації земельних покриттів чи виявлення змін на поверхні Землі). Водночас зібрати таку кількість знімків буває проблематично. GAN моделі здатні синтезувати нові супутникові фото, які будуть максимально наближені до реальних, а це допомагає ефективно збільшувати обсяги наборів даних.

Основним недоліком традиційних GAN моделей є те, що результат, який вони генерують, не завжди можна передбачити. Цю проблему було усунуто в умовних мережах [3], які дозволяють контролювати процес генерації за допомогою додаткового сигналу – умови, якою може бути вектор, що містить певну інформацію про бажаний результат, наприклад, приналежність до конкретного класу, стиль чи набір візуальних атрибутів. Попри це найбільш інтерактивним для користувача методом управління процесом генерації зображення вважається керування за допомогою довільного текстового опису. Цей підхід передбачає накладання одночасно декількох умов та опис очікуваного результату природною мовою.

Використання мовних моделей дозволяє перетворювати текст у вектор, який можна застосовувати для навчання GAN моделей. Саме тому дослідження текстових кодувальників, що використовують сучасні мовні нейронні мережі, є ключовим напрямком у вдосконаленні генеративних мереж для створення супутникових зображень. Ці кодувальники здатні ефективно перетворювати складні текстові описи користувачів у високорівневі векторні представлення, які точно відображають необхідні характеристики зображень. Це дозволяє створювати більш комплексні та багатовимірні супутникові фотознімки, які враховують різні аспекти реального світу, забезпечуючи високу точність і релевантність результатів для різних галузей застосування.

Аналіз досліджень та публікацій

Залежно від поставленого завдання структура дискримінатора та генератора GAN моделі може видозмінюватися, оскільки неправильне поєднання векторів ознак із представленнями зображень призводить до зменшення точності та стійкості моделі [3].

У статті [4] представлено мережу StackGAN, у якій використано двоетапну архітектуру для генерації високоякісних зображень із текстових описів. Перший етап генерує базове зображення з низькою роздільною здатністю, а другий покращує деталі та якість. Такий підхід дозволяє досягти кращої деталізації порівняно з однорівневими моделями. Проте StackGAN стикається з проблемою синхронізації між двома етапами генерації, що може призводити до нестабільності під час навчання. Крім того, модель не завжди ефективно аналізує семантичні ознаки вхідних даних, а це обмежує її здатність генерувати зображення з високим ступенем відповідності до текстових описів.

У роботі [5] запропоновано використання механізму уваги для покращення відповідності між реченнями та згенерованими зображеннями. AttnGAN дозволяє моделі фокусуватися на релевантних частинах тексту під час генерації кожної частини зображення, що сприяє більш точному відтворенню деталей. Однак впровадження механізму уваги значно ускладнює архітектуру моделі і збільшує вимоги до обчислювальних ресурсів та часу навчання. Крім того, AttnGAN може бути недостатньо ефективним при роботі з великими та складними текстовими описами, характерними для супутникових зображень, де потрібна висока точність та деталізація.

У статті [6] представлено генеративну мережу DM-GAN, яка використовує динамічну пам'ять для покращення процесу генерації зображень. Нейронна мережа зберігає та оновлює інформацію про текстові описи впродовж усіх етапів генерації, що покращує узгодженість між текстом і зображенням. Проте інтеграція динамічної пам'яті у архітектуру GAN ускладнює процес навчання та вимагає додаткових механізмів для управління пам'яттю. Це може призводити до проблем зі стабільністю тренування та збільшує час, який необхідний для досягнення оптимальної якості зображень, оскільки процес навчання потребує великої кількості епох.

Після анонсу моделі BERT [7] було декілька успішних спроб адаптації кодувальника тексту на його основі в GAN моделі [8, 9, 10], але вони мають ряд суттєвих обмежень. Стандартна BERT модель генерує контекстуальні представлення на рівні tokenів, що може бути недостатньо ефективним для вирішення завдань, коли необхідно згенерувати зображення за реченнями, що містять велику кількість слів. Це також негативно впливає на здатність моделей точно передавати складні семантичні взаємозв'язки між текстовими описами та зображеннями, особливо у випадку супутникових знімків, інструкції для створення яких можуть містити детальні та специфічні описи. Для подолання цих обмежень було розроблено модель Sentence-BERT (SBERT), яка спеціально оптимізована для отримання семантично узгоджених векторних представлень речень [11]. Архітектура SBERT дозволяє враховувати контекст цілого речення, що забезпечує «наближення» семантично схожих текстів у векторному просторі. Модель здатна аналізувати семантичне значення текстових описів, що є критично важливим для коректної інтерпретації інструкцій користувача та генерації складних зображень, таких як супутникові фотознімки.

Формулювання цілей статті

Метою роботи є генерація зображень на основі тексту за допомогою різних типів GAN поєднаних з трансформером SBERT, зокрема:

- проведення оцінки якості згенерованих зображень та визначення найпродуктивніших

нейронних мереж, які створюють зображення, що найбільш близькі до реальних супутникових фото.

- дослідження впливу кількості епох навчання на точність генерації зображень;

Описані в цій статті методи та результати можуть використовуватися для покращення точності генерації зображень тексту за допомогою різних GAN моделей (LSGAN, DCGAN, WGAN, WGAN-GP) поєднаних з трансформером SBERT.

Виклад основного матеріалу

На відміну від традиційного BERT, який створює контекстуалізовані вектори для кожного токена, SBERT генерує фіксовані векторні представлення для всього речення. Вектори, отримані за допомогою SBERT, мають фіксовану довжину, яка залежить від конкретної архітектури моделі. Наприклад, для моделі all-mpnet-base-v2, яка використовувалася для тренування мереж, довжина вектору складає 768, що дозволяє захоплювати складні семантичні відносини між різними текстовими описами.

Для того, щоб навчати генератор G нейронної моделі GAN, матриці шуму z та представлення тексту e об'єднувалися в одну матрицю x_G , яка використовувалася для генерації вихідного зображення $\hat{x}_G$:

$$x_G = \begin{bmatrix} z \\ e \end{bmatrix} \in \mathbb{R}^{k+d}, \quad (1)$$

$$\hat{x}_G = G(x_G), \quad (2)$$

де k – довжина вектору довільних значень z ;

d – довжина векторного представлення тексту e .

У дискримінаторі D здійснювалася проєкція тексту e в просторовий формат (матрицю):

$$f(e) = Ve + b; f(e) \in \mathbb{R}^{H \times W}, \quad (3)$$

де V – матриця ваг;

b – вектор зміщення, що додається до проєкції для забезпечення гнучкості моделі;

H – висота зображення;

W – ширина зображення.

Використовуючи матричне представлення оригінального зображення x та $f(e)$, дискримінатор прогнозує, чи є зображення реальним, чи згенерованим:

$$\hat{y}_D = D \left(\begin{bmatrix} x \\ f(e) \end{bmatrix} \right), \quad (4)$$

де $\hat{y}_D$ – прогноз дискримінатора щодо того, чи є зображення реальним ($\hat{y}_D = 1$) або згенерованим ($\hat{y}_D = 0$).

Для того, щоб визначити як функції втрат та архітектура нейронних мереж впливають на якість згенерованих супутникових зображень, для проведення експериментів застосовувалися DCGAN [12], LSGAN [13], WGAN [14] та WGAN-GP [15].

DCGAN використовує згорткові нейронні мережі та стандартну мінімаксну функцію втрат [12]:

$$\min_G \max_D L(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log 1 - D(G(z))], \quad (5)$$

де $D(x)$ – дискримінатор, який оцінює, чи вихідне зображення x є реальним, а не згенерованим;

$G(z)$ – генератор, який перетворює шум у зображення.

На відміну від традиційних GAN, у LSGAN використовується функція втрат на основі найменших квадратів [13]:

$$\min_G L(D) = \frac{1}{2} \mathbb{E}_{x \sim p_{data}(x)} [(D(x) - b)^2] + \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - a)^2] \quad (6)$$

$$\min_G L(G) = \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - c)^2], \quad (7)$$

де a – цільове значення для фейкових даних (зазвичай 0);

b – цільове значення для реальних даних (зазвичай 1);

c – цільове значення, якого генератор прагне досягти для фейкових даних (зазвичай 1).

Генеративні мережі з архітектурою LSGAN часто генерують більш реалістичні та варіативні зображення ніж традиційні GAN.

Із метою стабілізації процесу навчання у 2017 році було запропоновано WGAN, який здобув популярність завдяки тому, що значення функції втрат критика (8) та генератора (9) більш інформативно відображають відмінності між реальними та згенерованими розподілами даних [14].

$$L_D = \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] - \mathbb{E}_{z \sim \mathbb{P}_z} [D(G(z))] \quad (8)$$

$$L_G = -\mathbb{E}_{z \sim \mathbb{P}_z} [D(G(z))], \quad (9)$$

де $\mathbb{P}_r$ – розподіл реальних даних;

$\mathbb{P}_z$ – розподіл згенерованих даних.

Однією із проблем WGAN є тривалий час навчання. Для того, щоб вирішити цю проблему, було запропоновано WGAN-GP, яка використовує градієнтний штраф. Він додається до функції втрат дискримінатора та стимулює градієнти критика мати норму близьку до 1 на випадково взятих точках між реальними та згенерованими даними [15]:

$$L_D = \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] - \mathbb{E}_{z \sim \mathbb{P}_z} [D(G(z))] + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\| - 1)^2], \quad (10)$$

де λ – параметр, який контролює силу штрафу;

$\mathbb{P}_{\hat{x}}$ – розподіл інтерпольованих точок між реальними та згенерованими даними.

Для навчання GAN моделей використовувався набір даних RSICD [16], який містить 10921

супутникове зображення, кожне з яких супроводжується 5 унікальними реченнями, які описують різні аспекти та характеристики зображень, такі як тип місцевості, об'єкти, погодні умови та інші важливі деталі. Усі зображення для тренування та валідації моделей були зібрані з відкритих джерел і не містять інформації, яка може порушувати конфіденційність даних.

Навчання здійснювалося за допомогою фреймворку PyTorch [17], а для оптимізації параметрів моделі застосовувався оптимізатор Адама [18] із кроком навчання $1e-4$. Для того, щоб уніфікувати зображення, перед тренуванням вони зменшувалися до розміру 64x64 пікселів, а кожен із RGB каналів зображення проходив нормалізацію. Розмір одного пакету навчання був обмежений до 64. Набір даних для тестування містив 100 зображень, довільно обраних із RSICD.

Для того, щоб протестувати точність нейронних мереж, використовувалася FID метрика [19], яка здатна кількісно оцінювати реалістичність та різноманітність створених зображень, а кількість епох змінювалася між дослідженнями, щоб визначити оптимальну кількість епох для навчання різних типів GAN моделей. У табл. 1 наведено результати тестування цих моделей із різною кількістю епох.

Таблиця 1

FID оцінки GAN моделей

GAN модель	Кількість епох N			
	N=50	N=100	N=150	N=200
DCGAN	48,6	41,4	35,8	33,2
LSGAN	53,7	47,2	42,4	39,1
WGAN	61,3	55,8	50,1	47,6
WGAN-GP	44,2	38,3	33,7	33,1

Для того, щоб оцінити структурну подібність згенерованих даних, використовувалася SSIM метрика [20], яка застосовується для оцінки схожості між двома зображеннями шляхом порівняння їхньої яскравості, контрасту та структурних особливостей. У табл. 2 наведено результати тестування GAN-моделей із різною кількістю епох за показником SSIM.

Таблиця 2

SSIM оцінки GAN моделей

GAN модель	Кількість епох N			
	N=50	N=100	N=150	N=200
DCGAN	0,63	0,67	0,72	0,73
LSGAN	0,59	0,63	0,67	0,69
WGAN	0,55	0,59	0,64	0,67
WGAN-GP	0,65	0,70	0,74	0,76

Результати демонструють, що збільшення кількості епох значно покращує точність GAN моделей, а при невеликій кількості вона залишається незадовільною. WGAN-GP модель показала найкращі результати, а значення FID=33,1 та SSIM=0,76 свідчать про високу реалістичність і структурну подібність згенерованих зображень.

Висновки

У роботі було досліджено доцільність інтеграції SBERT у структуру GAN моделей для створення супутникових зображень на основі тексту. Аналізуючи результати, можна побачити, що тип нейронних мереж GAN значно впливає на точність перетворення тексту в зображення. Найпродуктивніше себе зарекомендувала нейронна мережа WGAN-GP. Зображення, які були створені за її допомогою, є чіткими, краще відтворюють текстури реальних супутникових зображень та містять низьку кількість артефактів. Це демонструє, що використання кодувальника тексту на основі SBERT дозволяє ефективно отримувати семантичну інформацію із текстових описів та перетворювати її у відповідні векторні представлення для навчання генератора та дискримінатора. Результати експериментів вказують напрям для подальшого вдосконалення подібних гібридних моделей. Зокрема можлива інтеграція інших глибоких чи спеціалізованих мовних моделей задля аналізу семантичної інформації, а також адаптація архітектури генеративних мереж, наприклад, модифікація шарів пулінгу (агрегації) між згортковими шарами [21] для урахування специфіки різних типів зображень.

Під час проведення експериментів було виявлено, що для збільшення точності моделей критично необхідно збільшувати кількість епох навчання моделей. Це пояснюється тим, що SBERT забезпечує якісне семантичне представлення текстових описів, проте для досягнення більш точної відповідності між текстом і зображенням модель потребує більшої кількості ітерацій. На початкових етапах навчання генератор не встигає опрацювати всі аспекти семантики, закодованої SBERT, а це призводить до менш

якісних та деталізованих результатів. Проте зі збільшенням кількості епох навчання генератор здатний краще відтворювати структуру та текстури об'єктів, за векторним представленням тексту.

Невеликі значення оцінок FID та SSIM, отримані для моделей WGAN та LSGAN, свідчать про те, що запропонований підхід до інтеграції кодувальника тексту SBERT у структуру GAN має значний потенціал для подальших досліджень. Одним із перспективних напрямів є детальніше дослідження морфології прихованих нейронних шарів GAN, які відповідають за перетворення векторних представлень тексту, та визначення того, наскільки їхня конфігурація впливає на якість і точність роботи генеративної моделі.

Результати дослідження показують, що мовна модель SBERT може ефективно використовуватися як кодувальник тексту для GAN моделей завдяки здатності SBERT отримувати семантичні представлення текстової інформації, що дозволяє точно передавати зміст текстових описів у векторному форматі. Це особливо важливо для користувачів систем генерування зображень, оскільки сприяє створенню контенту, який відповідає їхнім очікуванням та потребам.

Література

1. Goodfellow I. Generative adversarial networks / I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio // *Communications of the ACM*. – 2020. – Vol. 63, No. 11. – P. 139–144.
2. Mansourifar H. GAN-Based Satellite Imaging: A Survey on Techniques and Applications / H. Mansourifar, A. Moskovitz, B. Klingensmith, D. Mintas, S. J. Simske // *IEEE Access*. – 2022. – Vol. 10. – P. 118123–118140. – DOI: <https://doi.org/10.1109/ACCESS.2022.3221123>.
3. Mirza M. Conditional generative adversarial nets / M. Mirza // *arXiv preprint arXiv:1411.1784*. – 2014. – URL: <https://arxiv.org/abs/1411.1784>.
4. Zhang H. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks / H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D. N. Metaxas // *Proceedings of the IEEE International Conference on Computer Vision*. – 2017. – P. 5907–5915.
5. Xu T. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks / T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, X. He // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. – 2018. – P. 1316–1324.
6. Zhu M. DM-GAN: Dynamic memory generative adversarial networks for text-to-image synthesis / M. Zhu, P. Pan, W. Chen, Y. Yang // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. – 2019. – P. 5802–5810.
7. Devlin J. BERT: pre-training of deep bidirectional transformers for language understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*. – 2019. – Vol. 1. – P. 4171–4186.
8. Croce D. GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples / D. Croce, G. Castellucci, R. Basili // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. – 2020. – P. 2114–2119.
9. Na S. Realistic Image Generation from Text by Using BERT-Based Embedding / S. Na, M. Do, K. Yu, J. Kim // *Electronics*. – 2022. – Vol. 11. – DOI: <https://doi.org/10.3390/electronics11050764>.
10. Tsue T. Cycle text-to-image GAN with BERT / T. Tsue, S. Sen, J. Li // *arXiv preprint arXiv:2003.12137*. – 2020. – URL: <https://arxiv.org/abs/2003.12137>.
11. Reimers N. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks / N. Reimers, I. Gurevych // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. – 2019. – P. 3982–3992. – DOI: <https://doi.org/10.18653/v1/D19-1410>.
12. Radford A. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks / A. Radford, L. Metz, S. Chintala // *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. – 2016. – URL: <http://arxiv.org/abs/1511.06434>.
13. Mao X. Least squares generative adversarial networks / X. Mao, Q. Li, H. Xie, R. YK Lau, Z. Wang, S. P. Smolley // *Proceedings of the IEEE International Conference on Computer Vision*. – 2017. – P. 2794–2802.
14. Adler J., Lunz S. Banach Wasserstein GAN / J. Adler, S. Lunz // *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. – 2018. – P. 6755–6764.
15. Gulrajani I. Improved Training of Wasserstein GAN / I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville // *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, NeurIPS 2017, December 4-9, 2017, Long Beach, CA, USA*. – 2017. – P. 5767–5777.
16. Lu X. Exploring models and data for remote sensing image caption generation / X. Lu, B. Wang, X. Zheng, X. Li // *IEEE Transactions on Geoscience and Remote Sensing*. – 2017. – Vol. 56, № 4. – P. 2183–2195.

17. Paszke A. PyTorch: An Imperative Style, High-Performance Deep Learning Library / A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury // *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada.* – 2019. – P. 8024–8035.
18. Kingma D. P. Adam: A Method for Stochastic Optimization / D. P. Kingma, J. Ba // *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.* – 2015.
19. Heusel M. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium / M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter // *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, NeurIPS 2017, December 4-9, 2017, Long Beach, CA, USA.* – 2017. – P. 6626–6637.
20. Wang Z. Image quality assessment: from error visibility to structural similarity / Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli // *IEEE Transactions on Image Processing.* – 2004. – Vol. 13, № 4. – P. 600–612.
21. Peleshchak R. Morphology of convolutional neural network with diagonalized pooling / R. Peleshchak, V. Lytvyn, O. Mediakov, I. Peleshchak // *International Conference on Modelling and Development of Intelligent Systems.* – 2022. – P. 161–172.

References

1. Goodfellow I. Generative adversarial networks / I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio // *Communications of the ACM.* – 2020. – Vol. 63, No. 11. – P. 139–144.
2. Mansourifar H. GAN-Based Satellite Imaging: A Survey on Techniques and Applications / H. Mansourifar, A. Moskovitz, B. Klingensmith, D. Mintas, S. J. Simske // *IEEE Access.* – 2022. – Vol. 10. – P. 118123–118140. – DOI: <https://doi.org/10.1109/ACCESS.2022.3221123>.
3. Mirza M. Conditional generative adversarial nets / M. Mirza // *arXiv preprint arXiv:1411.1784.* – 2014. – URL: <https://arxiv.org/abs/1411.1784>.
4. Zhang H. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks / H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D. N. Metaxas // *Proceedings of the IEEE International Conference on Computer Vision.* – 2017. – P. 5907–5915.
5. Xu T. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks / T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, X. He // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* – 2018. – P. 1316–1324.
6. Zhu M. DM-GAN: Dynamic memory generative adversarial networks for text-to-image synthesis / M. Zhu, P. Pan, W. Chen, Y. Yang // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* – 2019. – P. 5802–5810.
7. Devlin J. BERT: pre-training of deep bidirectional transformers for language understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019.* – 2019. – Vol. 1. – P. 4171–4186.
8. Croce D. GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples / D. Croce, G. Castellucci, R. Basili // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* – 2020. – P. 2114–2119.
9. Na S. Realistic Image Generation from Text by Using BERT-Based Embedding / S. Na, M. Do, K. Yu, J. Kim // *Electronics.* – 2022. – Vol. 11. – DOI: <https://doi.org/10.3390/electronics11050764>.
10. Tsue T. Cycle text-to-image GAN with BERT / T. Tsue, S. Sen, J. Li // *arXiv preprint arXiv:2003.12137.* – 2020. – URL: <https://arxiv.org/abs/2003.12137>.
11. Reimers N. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks / N. Reimers, I. Gurevych // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).* – 2019. – P. 3982–3992. – DOI: <https://doi.org/10.18653/v1/D19-1410>.
12. Radford A. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks / A. Radford, L. Metz, S. Chintala // *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings.* – 2016. – URL: <http://arxiv.org/abs/1511.06434>.
13. Mao X. Least squares generative adversarial networks / X. Mao, Q. Li, H. Xie, R. YK Lau, Z. Wang, S. P. Smolley // *Proceedings of the IEEE International Conference on Computer Vision.* – 2017. – P. 2794–2802.
14. Adler J., Lunz S. Banach Wasserstein GAN / J. Adler, S. Lunz // *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada.* – 2018. – P. 6755–6764.
15. Gulrajani I. Improved Training of Wasserstein GAN / I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville // *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, NeurIPS 2017, December 4-9, 2017, Long Beach, CA, USA.* – 2017. – P. 5767–5777.
16. Lu X. Exploring models and data for remote sensing image caption generation / X. Lu, B. Wang, X. Zheng, X. Li // *IEEE Transactions on Geoscience and Remote Sensing.* – 2017. – Vol. 56, № 4. – P. 2183–2195.
17. Paszke A. PyTorch: An Imperative Style, High-Performance Deep Learning Library / A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury // *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada.* – 2019. – P. 8024–8035.
18. Kingma D. P. Adam: A Method for Stochastic Optimization / D. P. Kingma, J. Ba // *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.* – 2015.
19. Heusel M. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium / M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter // *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, NeurIPS 2017, December 4-9, 2017, Long Beach, CA, USA.* – 2017. – P. 6626–6637.
20. Wang Z. Image quality assessment: from error visibility to structural similarity / Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli // *IEEE Transactions on Image Processing.* – 2004. – Vol. 13, № 4. – P. 600–612.
21. Peleshchak R. Morphology of convolutional neural network with diagonalized pooling / R. Peleshchak, V. Lytvyn, O. Mediakov, I. Peleshchak // *International Conference on Modelling and Development of Intelligent Systems.* – 2022. – P. 161–172.