

КРИВЕНЧУК ЮРІЙ

Національний університет "Львівська політехніка"

<https://orcid.org/0000-0002-2504-5833>e-mail: [Yurii.P.Kryvenchuk@lpnu.ua](mailto:Yurii.P.Kryvenchuk@lpnu.ua)

ЧАБАН СОФІЯ

Національний університет "Львівська політехніка"

e-mail: [sofia.chaban.knm.2022@lpnu.ua](mailto:sofia.chaban.knm.2022@lpnu.ua)

## НАВЧАННЯ ГЕНЕРАТИВНО-ЗМАГАЛЬНИХ МОДЕЛЕЙ ПЕРЕКЛАДУ ЗОБРАЖЕНЬ В УМОВАХ ОБМЕЖЕНОЇ КІЛЬКОСТІ ДАНИХ

Сучасний світ відзначається стрімким розвитком інформаційних технологій та широким застосуванням штучного інтелекту у багатьох сферах життя. Однією з ключових областей, де штучний інтелект демонструє вражаючий потенціал, є машинне та глибоке навчання. Зокрема, генеративно-змагальні мережі (GANs) представляють собою потужний інструмент для створення реалістичних зображень, а також для перекладу між різними мовами і модальностями. Найбільш вражаючою є модель CycleGAN, оскільки вона є потужним інструментом для перекладу зображень між різними доменами без необхідності парних даних для навчання. Вона відзначається здатністю до перетворення зображень одного типу в інший з вражаючою точністю, забезпечуючи циклічну консистентність між доменами. Однак, однією з головних обмежень застосування GANs є потреба у великій кількості даних для їх навчання. При навчанні на обмежених даних генеративно-змагальні моделі, такі як CycleGAN, можуть виявляти схильність до надмірної генералізації. Це означає, що модель може намагатися підлаштуватися під обмежену навчальну вибірку занадто сильно, що впливає на її здатність до адаптації до різноманітних вхідних даних. Модель може не мати достатнього розмаїття для відтворення різних аспектів вихідних зображень, що призводить до втрати деталей та візуальних артефактів у створених зображеннях. В умовах, коли існує обмежена кількість даних, розробка ефективних методів навчання генеративно-змагальних моделей стає надзвичайно актуальною задачею. Дана робота присвячена дослідженню та розробці таких методів навчання генеративно-змагальних моделей для вирішення завдань перекладу зображень в умовах обмеженої кількості доступних даних.

Ключові слова: генеративно-змагальні-мережі, аугментація даних, переклад зображень

KRYVENCHUK YURIY, CHABAN SOFIIA

Lviv Polytechnic National University

### TRAINING GENERATIVE ADVERSARIAL MODELS FOR IMAGE TRANSLATION UNDER CONDITIONS OF LIMITED DATA

In our rapidly evolving modern world, characterized by the swift progress of information technology and the widespread integration of artificial intelligence across diverse domains, the potential of artificial intelligence shines prominently, particularly in the realms of machine and deep learning. Within this context, Generative Adversarial Networks (GANs) emerge as formidable tools, adept at producing realistic images and facilitating seamless translations across different languages and modalities. Notably, the CycleGAN model distinguishes itself as a formidable instrument for image translation between disparate domains, dispensing with the necessity of paired training data. Its exceptional precision in metamorphosing images from one category to another underscores its ability to maintain cyclic consistency across domains. Nevertheless, a principal challenge looms in the application of GANs—namely, the imperative of a substantial volume of data for effective training. When these generative adversarial models, exemplified by CycleGAN, are trained on limited datasets, they may incline toward excessive generalization. This tendency implies that the model might overfit to the restricted training samples, thereby compromising its adaptability to diverse input data. Consequently, the model might lack the requisite diversity to faithfully reproduce various facets of the source images, resulting in a loss of detail and the emergence of visual artifacts in the generated images. In scenarios where data availability is constrained, the imperative of devising efficient training methodologies for generative adversarial models becomes all the more salient. The present work is dedicated to the comprehensive exploration and development of such training methodologies, with a specific focus on addressing the intricate challenges of image translation tasks within the confines of limited available data. Through innovative strategies, we aim to enhance the robustness and generalization capacity of GANs, facilitating their effective application in real-world scenarios characterized by data scarcity.

Keywords: Generative Adversarial Networks, data augmentation, image translation.

### Постановка проблеми

Розробка моделей перекладу зображень, які працюють на основі GAN, відкрила перед нами нові можливості в галузях мистецтва, розваг, моди та медичної діагностики. Проте головною проблемою у використанні всього потенціалу цих моделей є їх підвищена вимогливість до даних. Зазвичай приймається, що для навчання глибоких моделей нейронних мереж, таких як GAN, потрібні великі набори даних. Це стає серйозним обмеженням у багатьох практичних застосуваннях, де збір обширних та якісних даних є або неможливим, або фінансово необґрунтованим. В більшості випадків класичні методи аугментації даних не підходять, оскільки модель GAN вимагає високої різноманітності. При генерації високоякісних зображень, потребують великого різноманіття в тренувальних даних. Класичні методи аугментації можуть здійснювати обмежений змішування, обертання або зміщення, але вони не завжди здатні створювати достатню різноманітність для GANs. Також часто виникає проблема перенавчання генератора. При обмеженій кількості тренувальних даних класичні методи аугментації можуть призвести до перенавчання моделі. Наприклад, занадто інтенсивне зміщення, обертання або деформація може призвести до того, що модель

навчитися "запам'ятовувати" аугментовані зразки, а не створювати нові. У зв'язку з цим дослідження та впровадження спеціалізованих методів аугментації даних, призначених саме для задач тренування GANs в умовах обмеженої кількості даних є актуальною темою, чому і буде присвячена ця робота.

### **Аналіз останніх досліджень**

В роботі [1] автори дослідили проблему тренування генеративно-змагальних на малій кількості даних та запропонували метод, який називається Differentiable Augmentation (DiffAugment), який покращує ефективність використання даних для GANs шляхом застосування різних видів диференційованих аугментацій як до реальних, так і до згенерованих зображень. Експерименти показують стійку покращену продуктивність методу над різними архітектурами GAN і функціями втрат для умовного та безумовного генерування зображень. У дослідженні [2] автори пропонують механізм адаптивного аугментації дискримінатора, який значно стабілізує процес навчання при обмеженій кількості даних. Цей підхід не вимагає змін у функціях втрат або архітектурі мережі і може бути використаний як для навчання моделі "з нуля", так і для налаштування існуючої GAN на інших даних. У роботі [3] пропонується метод FreGAN, який підвищує увагу моделі до частот і сприяє синтезу високочастотних сигналів, що полегшує створення високоякісних зображень. Крім використання частотної інформації як для реальних, так і для згенерованих зображень, автори використовують інформацію про частоти реальних зображень як самонавчальний обмежувальний фактор, який знімає дисбаланс GAN і спонукає генератор створювати адекватні, а не випадкові частотні сигнали. У статті [4] дослідники використали метод ScoreMix ефективно збільшує різноманітність даних та зменшує проблему перенавчання. Більше того, його легко можна включити в існуючі моделі GAN з незначними модифікаціями. Експериментальні результати на численних завданнях демонструють значні покращення в роботі GAN-моделей, які використовують метод ScoreMix. В роботі [5] автори застосували стратегію, в якій навчання починається з безумовної GAN і поступово вводить класове умовлення в генератор і функцію цілі. Запропонований метод для навчання cGANs з обмеженими даними призводить до стабільного навчання і генерації високоякісних зображень завдяки ранньофазовій експлуатації спільної інформації між класами. У статті [6] пропонується нова стратегія, яку називають Adaptive Pseudo Augmentation (APA), щоб стимулювати здорову конкуренцію між генератором і дискримінатором. APA відзначається як альтернативний метод порівняно з існуючими підходами, які базуються на стандартних методах аугментації даних або регуляризації моделей.

В роботі [7] проведено обширне емпіричне дослідження, яке підтверджує, що запропонована стратегія навчання, названа DynamicD, покращує якість синтезу без додаткових обчислювальних витрат або цілей навчання. Розроблено два схеми адаптації потужності дискримінатора: i) при наявності достатньої кількості тренувальних даних дискримінатор користується поступовим збільшенням потужності навчання, і ii) при обмеженій кількості тренувальних даних поступове зменшення ширини шару допомагає уникнути перенавчання дискримінатора.

У статті [8] автори розробили підхід, в якому вони перепараметризують простір генерації як суміш моделі і навчають параметри цієї суміші разом з параметрами GAN. Ця здається простою модифікацією GAN, але виявляється дивовижно ефективною, дозволяючи отримати різноманітність у згенерованих зразках навіть при обмежених даних.

У роботі [9] пропонується механізм аугментації даних для поліпшення розміру, якості і різноманітності тренувального набору даних. Ця аугментація включає в себе різні перетворення, такі як обертання, зсуви, афінні перетворення і т. д. В роботі використовується відстань Фреше (Fretchet distance) як метрика для оцінки ефективності різних аугментаційних методів на тренувальному наборі даних.

У статті [10] запропоновано новий метод передачі знань для генеративних моделей на основі видобування знань, які найбільш корисні для конкретної цільової області, незалежно від того, чи вони походять від одного, чи від кількох попередньо навчених GANs. Для цього використовується мережа "miner", яка визначає, яка частина генеративного розподілу кожного попередньо навченого GAN найближча до цільової області.

### **Формулювання цілей статті**

**Метою роботи** є дослідження та розробка методів навчання генеративно-змагальних моделей для завдань перекладу зображень в умовах обмеженої кількості даних

### **Виклад основного матеріалу**

#### **Огляд набору даних**

Для цієї роботи було обрано "Horse2Zebra dataset". Набір даних "Horse2Zebra" — це колекція зображень, які використовуються в дослідженнях комп'ютерного зору, зокрема у сфері перекладу зображення в зображення та перетворення стилю. Цей набір даних призначений для завдань, які передбачають перетворення зображень коней на зображення зебр і навпаки. Він служить еталоном для навчання та оцінки моделей перекладу зображень, таких як Generative Adversarial Networks (GANs) і Variation Autoencoders (VAEs). Цей складається загалом з 1187 зображень коня та 1474 зебри.

#### **Етапи процесу перекладу зображень в умовах обмеженої кількості даних**

Процес перекладу зображень за допомогою генеративно-змагальних мереж можна поділити на такі основні етапи: збір даних, попередня обробка даних, вибір моделі, модифікація моделі для застосування методів аугментації даних, тренування моделі, оцінка моделі.

*Збір даних.* Перший крок – це зібрати набір даних з пар зображень, де кожна пара складається з

оригінального зображення та його відповідного перекладеного зображення. Наприклад, якщо стоїть задача перекладу зображення з літнього пейзажу в зимовий, то потрібно зібрати пари, де одне зображення – це літній пейзаж, а інше – зимовий пейзаж.

*Попередня обробка даних.* Для підготовки даних може знадобитися їх масштабування, обрізка або інша обробка зображень. Крім того, датасет має бути розділений на навчальний і тестовий набори для оцінки результатів.

*Вибір моделі.* Вибір архітектури і параметрів моделі є важливим завданням. Однією з найбільш популярних генеративно-змагальних мереж для задачі перекладу зображень є CycleGAN. Однією з основних переваг цієї архітектури є можливість навчати модель на непарних даних, тобто даних без явного співставлення між вхідним і вихідним доменами.

*Модифікація моделі для застосування методів аугментації даних.* Застосування класичних методів аугментації даних може призвести до того, що генератор навчиться створювати трансформовані зображення. Наприклад, аугментація шуму призводить до виникнення шумних результатів, навіть якщо його немає в наборі даних. Для вирішення цієї проблеми необхідно внести зміну до стандартної архітектури GAN, як зображено на рис. 1. В цій версії над реальними та фейковими зображеннями виконується однакова трансформація ( $T_k$ ) перед тим, як дискримінатор проводить класифікацію.

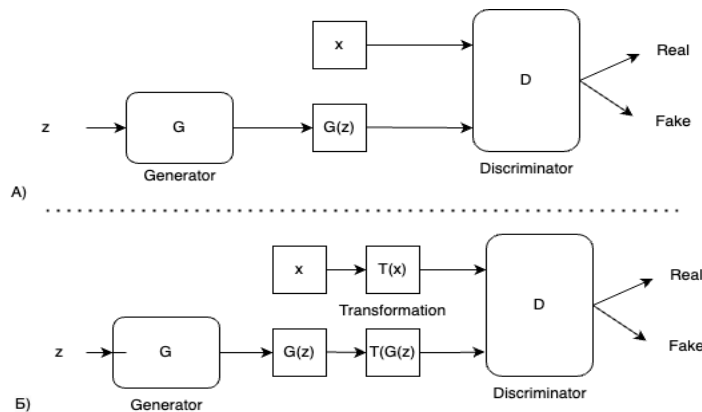


Рис. 1. Архітектури GAN: А) оригінальна архітектура GAN, Б) модифікована архітектура GAN для використання аугментації даних

*Тренування моделі.* Модель тренується на навчальних даних з метою мінімізації функції втрат, що визначає різницю між оригінальними і перекладеними зображеннями, та функцій втрат дискримінаторів. Під час тренування CycleGAN також застосовується циклічний переклад для покращення якості.

*Оцінка моделі.* Після завершення тренування моделі важливо оцінити її продуктивність. Це може включати в себе визначення якості перекладів на тестовому наборі, а також оцінку швидкодії моделі.

### Висновок

В результаті цієї роботи були досліджені та розроблені методи аугментації даних та вдосконалено стандартну архітектуру генеративно-змагальних мереж (GAN) з метою штучного збільшення кількості тренувальних даних. Ця розробка має велике практичне значення, оскільки вона дозволяє ефективно вирішити проблему обмеженості тренувальних даних, особливо актуальну в сферах, де зібрати велику кількість навчальних прикладів складно або дорого, таких як медицина.

Зростаючий інтерес до використання GAN і схожих методів у контексті обмежених даних підкреслює важливість подальших досліджень та розвитку цього напрямку. Розробка алгоритмів, які дозволяють ефективно використовувати обмежені дані, може принести користь у багатьох галузях, поліпшуючи точність, швидкість та доступність генеративних моделей.

### References

1. S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han, "Differentiable Augmentation for Data-Efficient GAN Training".
2. T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training Generative Adversarial Networks with Limited Data." arXiv, Oct. 07, 2020. <http://arxiv.org/abs/2006.06676>.
3. M. Yang, Z. Wang, Z. Chi, and Y. Zhang, "FreGAN: Exploiting Frequency Components for Training GANs under Limited Data".
4. J. Cao, M. Luo, J. Yu, M.-H. Yang, and R. He, "ScoreMix: A Scalable Augmentation Strategy for Training GANs With Limited Data," IEEE Transactions on Pattern Analysis and Machine Intelligence, вип. 45, p. 8920–8935, Jul. 2023, doi: 10.1109/TPAMI.2022.3231649.
5. M. Shahbazi, M. Danelljan, D. P. Paudel, and L. Van Gool, "Collapse by Conditioning: Training Class-conditional GANs with Limited Data." arXiv, Mar. 16, 2022. doi: 10.48550/arXiv.2201.06578.

6. L. Jiang, B. Dai, W. Wu, and C. C. Loy, “Deceive D: Adaptive Pseudo Augmentation for GAN Training with Limited Data,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, p. 21655–21667. <https://proceedings.neurips.cc/paper/2021/hash/b534ba68236ba543ae44b22bd110a1d6-Abstract.html>.
7. C. Yang, Y. Shen, Y. Xu, D. Zhao, B. Dai, and B. Zhou, “Improving GANs with A Dynamic Discriminator,” *Advances in Neural Information Processing Systems*, 35, p. 15093–15104, Dec. 2022.
8. S. Gurumurthy, R. Kiran Sarvadevabhatla, and R. Venkatesh Babu, “DeLiGAN : Generative Adversarial Networks for Diverse and Limited Data,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, p. 166–174. [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/Gurumurthy\\_DeLiGAN\\_Generative\\_CVPR\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/Gurumurthy_DeLiGAN_Generative_CVPR_2017_paper.html).
9. S. Lanka, G. Velingkar, R. Varadarajan, and M. Anand Kumar, “Image Augmentation Strategies to Train GANs with Limited Data,” in *Computer Vision and Machine Intelligence*, M. Tistarelli, S. R. Dubey, S. K. Singh, and X. Jiang, Eds., in *Lecture Notes in Networks and Systems*. Singapore: Springer Nature, 2023, p. 441–453. doi: 10.1007/978-981-19-7867-8\_35.
10. Y. Wang, A. Gonzalez-Garcia, D. Berga, L. Herranz, F. S. Khan, and J. van de Weijer, “MineGAN: Effective Knowledge Transfer From GANs to Target Domains With Few Images,” presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, p. 9332–9341. [https://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Wang\\_MineGAN\\_Effective\\_Knowledge\\_Transfer\\_From\\_GANs\\_to\\_Target\\_Domains\\_With\\_CVPR\\_2020\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2020/html/Wang_MineGAN_Effective_Knowledge_Transfer_From_GANs_to_Target_Domains_With_CVPR_2020_paper.html)