

БЕРДНИК ДАНИЛО

Національний університет "Львівська політехніка"

<https://orcid.org/0000-0002-8092-9228>e-mail: danylo.berdnyk.mknssh.2022@lpnu.ua

ПЕЛІШКО ДМИТРО

Національний університет "Львівська політехніка"

<https://orcid.org/0000-0003-4881-6933>e-mail: Dmytro.d.peleshko@lpnu.ua

ВІДНОВЛЕННЯ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ ГЕНЕРАТИВНИХ НЕЙРОННИХ МЕРЕЖ

В цьому дослідженні проводиться аналіз використання породжувальних нейронних мереж для реконструкції пошкоджених зображень. Для цього було використано систему з архітектурою контекстного кодувальника, який використовує змагальний процес навчання. Вид пошкодження зображень, який досліджувався у цій роботі – пропуски у зображеннях. Заповнення пропусків здійснюється на двох різних наборах даних за областями, а саме відновлення обличчя людей та відбитків пальців. Результати відновлення, які було отримано на тестувальних наборах даних було оцінено за допомогою таких метрик, як піксельні втрати Pixel Loss, пікове співвідношення сигналу до шуму PSNR та структурну подібність SSIM. Після проведення аналізу отриманих результатів, в межах дослідження, було запропоновано модифікації оригінальної архітектури генеративної мережі і використання втрати сприйняття під час тренування мережі, а також модифіковану, меншу за розмірами, архітектуру мережі, яка здатна досягати подібних результатів з меншим обсягом обчислень.

Ключові слова: породжувальні нейронні мережі, реконструкція зображень, контекстний кодувальник, змагальний процес навчання, пропуски у зображеннях, відновлення обличчя, відбитки пальців, піксельні втрати, PSNR, SSIM, модифікована архітектура, втрата сприйняття.

BERDNYK DANYLO, PELESHKO DMYTRO

Lviv Polytechnic National University

IMAGE RECONSTRUCTION USING GENERATIVE NEURAL NETWORKS

Nowadays, image reconstruction occupies an important place in image processing, solving the problem of recovering images from incomplete, noisy, or damaged data. Its applications cover various fields, including computer vision, medical imaging, etc. It is used to improve image quality, fill in gaps, remove unwanted elements, create new visual effects, etc. One of the main challenges of image reconstruction is the high dimensionality and complexity of real-world images, which often require sophisticated models and algorithms to determine their underlying structure and patterns. In addition, different types of image reconstruction tasks may have different constraints and objectives, such as preserving the original content or maintaining realism. There are many image reconstruction methods and techniques that can be divided into two categories: traditional methods and deep learning methods. Traditional methods rely on manually created features and mathematical models to formulate and solve the problem. Deep learning methods use neural networks to learn complex data structures and create realistic images.

This study provides analysis of the use of generative neural networks for the reconstruction of damaged images. For this purpose, we used a system with a contextual encoder architecture that uses an adversarial learning process. The type of image damage studied in this paper is inpainting. Inpainting is performed in two different domains, namely face and fingerprint recovery. The recovery results obtained on the test datasets were evaluated using such metrics as pixel loss, peak signal-to-noise ratio PSNR, and structural similarity SSIM. After analyzing the results, the study proposed modifications to the original architecture of the generative network. It was proposed to use the loss of perception during network training and a modified, smaller network architecture was proposed, which is able to achieve similar results with less computation.

Keywords: generative neural networks, image reconstruction, contextual encoder, adversarial learning process, inpainting, face reconstruction, fingerprints, pixel loss, PSNR, SSIM, modified architecture, perceptual loss

Постановка проблеми

У наш час реконструкція зображень займає важливе місце в обробці зображень, вирішуючи проблему відновлення зображень з неповних, зашумлених або пошкоджених даних. Її застосування охоплює різні сфери, включаючи комп'ютерний зір, медичну візуалізацію тощо. Вона використовується для покращення якості зображень, заповнення прогалів, видалення небажаних елементів, створення нових візуальних ефектів тощо. Однією з головних проблем реконструкції зображень є висока розмірність і складність зображень у реальному світі, які часто вимагають складних моделей і алгоритмів для визначення їхньої основної структури та патернів. Крім того, різні типи задач реконструкції зображень можуть мати різні обмеження та завдання, такі як збереження оригінального змісту або підтримання реалістичності. Існує багато методів і технік реконструкції зображень, які можна розділити на дві категорії: традиційні методи і методи глибокого навчання. Традиційні методи покладаються на створені вручну ознаки та математичні моделі для формулювання та вирішення проблеми. Методи глибокого навчання використовують нейронні мережі для вивчення складних структур даних і створення реалістичних зображень.

Одним з найпопулярніших методів реконструкції зображень є зафарбовування зображень, метою якого є заповнення відсутніх або пошкоджених областей на зображенні. Методи зафарбовування зображень зазвичай використовують інформацію з навколишніх пікселів або ділянок для інтерполяції або екстраполяції відсутніх значень. Деякі з традиційних методів зафарбовування зображень включають методи на основі

дифузії, методи на основі шаблонів та методи на основі розрідженого представлення. Однак ці методи мають певні обмеження, наприклад, вимагають ручного налаштування параметрів, чутливі до шуму та викидів, а також не здатні охопити високорівневу семантичну інформацію та загальну цілісність. Тому в останні роки методи глибокого навчання стали потужною альтернативою для розфарбовування зображень, особливо з розвитком генеративних нейронних мереж. Генеративні нейронні мережі – це клас нейронних мереж, які можуть навчитися генерувати реалістичні дані з деякого латентного представлення. Їх можна використовувати для розфарбовування зображень, вивчаючи відображення від пошкоджених або неповних даних до оригінальних або бажаних даних. Одним з найпопулярніших типів генеративних нейронних мереж для розфарбовування зображень є генеративні змагальні мережі (GAN).

Отже, обробка зображень залишається актуальною та важливою галуззю, адаптуючись до зростаючої складності та розмірності реальних зображень. Генеративні змагальні мережі та інші методи глибокого навчання оновлюють реконструкцію зображень, особливо в контексті зафарбовування, вирішуючи виклики з відновлення зображень з неповних, зашумлених або пошкоджених даних. Результати дослідження можуть бути корисними для розширення можливостей реконструкції зображень, зокрема у відновленні обличчя людей та відбитків пальців.

Аналіз останніх джерел

У роботі [1] запропоновано метод GANCS для реконструкції MPT-зображень, оснований на генеративних змагальних нейронних мережах (GAN), що дозволяє ефективно відновлювати зображення при обмеженому обсязі даних. Дослідження [2] присвячено методу UnfairGAN, який використовує покращену генеративну змагальну мережу для ефективного видалення крапель дощу на зображеннях, використовуючи високорівневу інформацію.

Автори [3] досліджують глибоку мережу з залишковим шаром для відновлення зображень у мікроскопії високої роздільності та локалізації молекул високої щільності. У [4] запропоновано покращення відновлення медичних зображень за допомогою поєднання GAN і Task-GAN, що включає патологічну класифікацію для як реальних, так і синтетичних зображень.

Проте застосування глибинних генеративних нейронних мереж для відновлення зображень потребує подальшого дослідження.

Формулювання цілей

Метою даного дослідження полягає у розробці архітектури та методів реконструкції зображень, застосовуючи генеративну нейронну мережу, для задачі заповнення пропусків у зображеннях. Також до мети входить розробка модифікованої оптимізованої версії системи для відновлення зображень.

Система для заповнення пропусків у зображеннях

Система для дослідження складається з моделі Context Encoder, яка використовує навчання за допомогою GAN (генеративно-змагальної мережі). Генеративно-змагальна нейронна мережа складається з двох нейронних мереж: генератора (декодер в нашому випадку) та дискримінатора.

Перед роботою з мережею, потрібно створити зображення з пошкодженою ділянкою. Для цього використовується накладання маски на зображення. Для відомих областей значення пікселів на масці дорівнює одиниці інакше - нулю. Нехай Img є вхідним зображенням розміром (c, p, m) , яке не має пропусків, а M є маскою однакового розміру з вхідним зображенням, тобто (c, p, m) . Маска M має значення 0 у тих позиціях, де потрібно створити пропуск, і значення 1 у тих позиціях, де потрібно залишити зображення без змін. Для того, щоб отримати відому область використовується операція поелементного множення на маску - $KnownArea = Img \odot M$. Відповідно, для отримання тої частини зображення, що була замаскована використовується обернені значення у масці M - $UnknownArea = Img \odot (1 - M)$.

Мережа кодера-декодера у системі, як видно з назви складається з кодувальника та декодувальника. Кодер приймає вхідне зображення I і зіставляє його з контекстним представленням C , яке фіксує важливу інформацію з відомих областей зображення - $C = E(I)$. Параметри кодера навчаються під час тренування, щоб мінімізувати різницю між передбачуваним контекстом та істинним контекстом, який отримується з маски (M). Декодер бере контекстне представлення C і генерує зафарбоване зображення I' з відсутніми або пошкодженими областями $I' = D(C)$.

Навчання моделі Context Encoder GAN проводилось з використанням таких функцій втрат: втрата при реконструкції, яка вимірює попіксельну різницю між згенерованим зафарбованим зображенням і еталонними (відомими) областями; змагальна втрата, яка вимірює, наскільки добре генератор може обдурити дискримінатор, щоб той класифікував його результати як справжні зображення; комбінація втрати при реконструкції та змагальної втрати.

Попіксельна втрата гарантує, що зафарбовані області відповідають оригінальному зображенню у відомих областях. Її часто визначають як іншу втрату MSE або L2, що кількісно визначає попіксельну різницю між зафарбованим зображенням і оригінальним зображенням, якщо обидва зображення доступні. Функція попіксельної втрати визначається за допомогою наступної формули:

$$L_{pix}(I', I) = \sum (\sum (I' - I)^2), \quad (1)$$

де I - частина оригінального зображення, а I' - відновлена частина зображення.

Враховуючи накладену маску на зображення попіксельну функцію запишемо наступним чином:

$$L_{pix}(I) = \|M \odot (I - CE((I - M) \odot I))\| \quad (2)$$

де CE – Context Encoder.

Змагальна втрата використовується для навчання генератора, в нашому випадку він також є декодером, створювати зафарбовані зображення, які є візуально переконливими і не відрізняються від реальних зображень. Для цього використовується кросентропія втрат:

$$L_{adv}(G, D) = -[\log(D(I_{real})) + \log(I - D(I_{fake}))] \quad (3)$$

де G - мережа-генератор, D - мережа-дискримінатор, I_{real} - реальне зображення, вибране з навчального набору даних. Завдання дискримінатора - класифікувати це зображення як справжнє, I_{fake} - зафарбоване зображення, згенероване генератором.

Враховуючи накладену маску на зображення змагальну функцію запишемо наступним чином:

$$L_{adv} = \max_D = E_{I_x} [\log(D(I)) + \log(I - D(CE((I - M) \odot I)))] \quad (4)$$

Перцептивна використовує особливості зображень, а не пікселі, для порівняння. Перцептивна втрата обчислює середньоквадратичну помилку між особливостями двох зображень на певному шарі попередньо навченої моделі, наприклад, у цьому дослідженні використовується архітектура VGG19. Перцептивна функція втрат визначається наступною формулою:

$$L_{perc}(I_1, I_2) = \frac{1}{N} \sum_{i=1}^N (v(I_1)_i - v(I_2)_i)^2 \quad (5)$$

де I_1 та I_2 це два зображення, які потрібно порівняти, V - це попередньо навчена модель, яка приймає на вхід зображення і видає його особливості на певному шарі, v - значення особливостей для зображення I на позиції i і N – це кількість особливостей на шарі моделі V.

В дослідженні використовувалось комбінація вищеписаних втрат:

$$L = \lambda_{pix} L_{pix} + \lambda_{adv} L_{adv} + \lambda_{perc} L_{perc} \quad (6)$$

Для оцінки отриманої мережі в результаті тренування, на тестувальній вибірці було використано такі метрики оцінок як PSNR та SSIM.

PSNR – це метрика, яка кількісно оцінює якість зображення, вимірюючи відношення максимально можливої потужності сигналу до потужності спотворюючого шуму.

$$PSNR = 10 * \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (7)$$

де MAX – максимально можливе значення пікселя. MSE – це середньоквадратична похибка між відповідними пікселями оригінального та реконструйованого зображень.

SSIM – це метрика, яка оцінює схожість між двома зображеннями, беручи до уваги яскравість, контраст і структуру. Вона порівнює локальні патерни інтенсивності пікселів, а не лише окремі значення пікселів. SSIM базується на трьох компонентах: яскравість, контрастність та структура.

$$SSIM(I_{orig}, I_{reconstr}) = \frac{(2\mu_{lorig} \mu_{lreconstr} + c_1)(2\sigma_{loriglreconstr} + c_2)}{(\mu_{lorig}^2 + \mu_{lreconstr}^2 + c_1)(\sigma_{lorig}^2 + \sigma_{lreconstr}^2 + c_2)} \quad (8)$$

де μ - середнє значення зображення, σ - стандартне відхилення зображення, $\sigma_{loriglreconstr}$ - коваріація оригінального та відновленого зображень, c_1 та c_2 - константи для стабілізації ділення з маленьким знаменником..

Результати експериментів

Для експериментів аналізу було обрано два набори даних: CelebA та SOCOFing. CelebA – це великий набір даних, що охоплює понад 200 000 зображень розміром 128x128 пікселів та 3 канала для кольорів. SOCOFing – це біометричний набір даних, що охоплює 6 000 чорно білих зображень відбитків пальців різних розмірів, які були зведені до спільного розміру 128x128 пікселів. Для кожного з наборів даних було накладено маску для тренування контекстного кодувальника розміром 64 x 64 пікселів.

Для початку експерименти були проведені для різних комбінацій попередньо описаних наборів даних на оригінальній архітектурі CE без та з додаванням втрати сприйняття (рис. 1, таблиця 1).

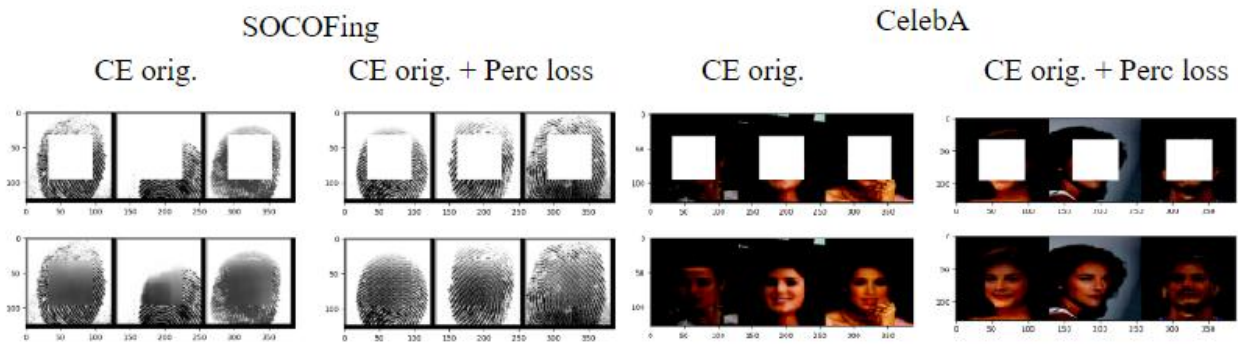


Рис. 1. Візуальні результати роботи оригінальної моделі CE з перцептивною втратою та без на обох наборах даних

Таблиця 1

Результати роботи оригінальної моделі CE з перцептивною втратою та без на обох наборах даних

Архітектура моделі	Набір даних	Pixel Loss	PSNR	SSIM
Context Encoder orig	CelebA	0.174	17.8	0.656
Context Encoder orig + Perc loss	CelebA	0.22	16.5	0.537
Context Encoder orig	SOCOFing	0.174	18.2	0.608
Context Encoder orig + Perc loss	SOCOFing	0.201	17.4	0.507

Далі, у рамках цього дослідження, оригінальну архітектуру контекстного кодувальника було модифіковано шляхом змінення саме архітектури частини генератора мережі. Така архітектура містить в 23.43 рази менше параметрів ніж оригінальна модель: 1723827, в той час, як оригінальна має 40401187 параметрів. Також така мережа генератора показує близько в 10 разів зменшення інференсу в часі на CPU: оригінальна модель - 1.01 сек.; модифікована - 0.1 сек.

Для модифікованої архітектури, за аналогією з оригінальним Context Encoder, було проведено експерименти з наборами даних CelebA та SOCOFing та з використанням втрати сприйняття (рис. 2, таблиця 2).

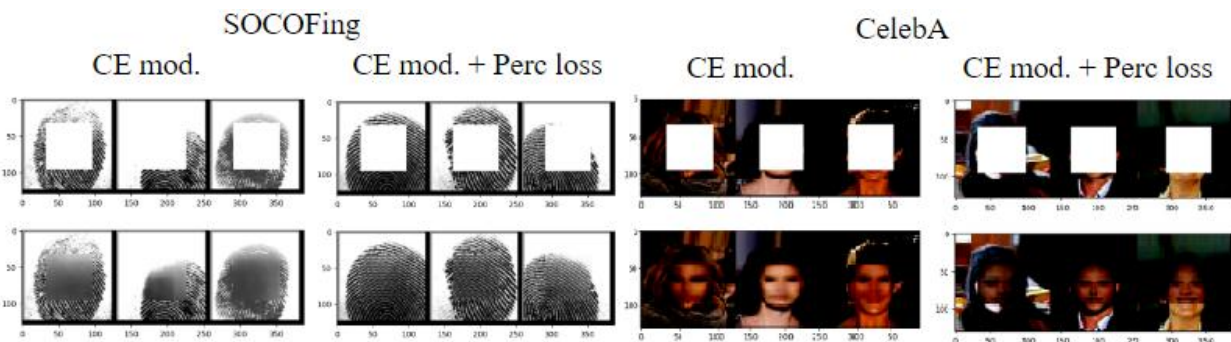


Рис. 2. Візуальні результати роботи модифікованої моделі CE з перцептивною втратою та без на обох наборах даних

Таблиця 2

Результати роботи модифікованої моделі CE з перцептивною втратою та без на обох наборах даних

Архітектура моделі	Набір даних	Pixel Loss	PSNR	SSIM
Context Encoder mod	CelebA	0.191	17.0	0.604
Context Encoder mod + Perc loss	CelebA	0.245	15.8	0.477
Context Encoder mod	SOCOFing	0.178	18.0	0.595
Context Encoder mod + Perc loss	SOCOFing	0.202	17.4	0.522

Висновки

У проведеній дослідницькій роботі було проведено детальний аналіз предметної області, описано процес діяльності та структуру даних, а також розглянуто існуючі аналоги з їхніми перевагами та недоліками. Вибір технологій та методів для дослідження був обґрунтований з урахуванням цього аналізу.

Також, за цим аналізом було сформульовано задачу роботи разом з підзадачами, які потрібно виконати для досягнення головної цілі дослідження. У другому розділі було розроблено систему дослідження та побудовано функціональну та математичну модель. В цій системі використовувалася Context Encoder модель разом із GAN. Ця модель включає кодер-декодер модель для врахування контексту зображення для аналізу зображень. Також система складається з даних облич людей та відбитків пальців, для яких, в межах системи накладається маска, яка симулює пошкодження зображень. Для кожного з елементів системи було описано математичні формулювання. У третьому розділі було описано програмне та технічне забезпечення, включаючи засоби розробки, програмну реалізацію для дослідження, а також результати експериментів. В цьому розділі було проведено експерименти для кожного з обраних наборів даних та різних модифікацій оригінальної мережі, таких, як застосування перцептивної втрати при тренування моделі і оптимізація оригінальної мережі шляхом зміни її структури. Показано, що використання перцептивної втрати в моделі Context Encoder покращує якість відтворення зображень, зокрема відбитків пальців. Також було продемонстровано, що модифікована архітектура мережі, яка має менше параметрів та швидший час інференсу, може досягати подібних результатів до оригінальної моделі для набору даних з відбитками пальців. Проте, для набору даних з обличчями результати зменшеної моделі є гіршими суттєво, але, використання перцептивної втрати, в свою чергу, покращує результати модифікованої моделі для цього набору даних. Отже, дослідження показало, що використання перцептивної втрати та модифікованих архітектур може покращити якість відтворення зображень у системі контекстного кодувальника, що має практичний потенціал у різних областях, включаючи обробку відбитків пальців.

References

1. Mardani M., Gong E., Cheng J. Y. Deep generative adversarial neural networks for compressive sensing mri. *IEEE Transactions on Medical Imaging*. 2019. Vol. 38, No. 1. P. 167–179.
2. Nguyen D. M., Le T. P., Vo D. M., Lee S.-W. UnfairGAN: an enhanced generative adversarial network for raindrop removal from a single image. *Expert Systems with Applications*. 2022. Vol. 210.
3. Image reconstruction with a deep convolutional neural network in high-density super-resolution microscopy.
4. Ouyang J., Wang G., Gong E. Task-gan: improving generative adversarial network for image reconstruction. Cham: Springer International Publishing, 2019.