

УДК 004.032.26:004.85]:004.93'1

DOI: 10.31673/2412-9070.2024.022729

О. Ю. КОТУЛ, студент;

О. В. СЕНЬКОВ, канд. техн. наук, доцент;

О. В. ЖИДКА, аспірантка,

Державний університет інформаційно-комунікаційних технологій, Київ

## АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ ТА СИНТЕЗУ ЗОБРАЖЕНЬ ДЛЯ РОЗШИРЕННЯ НАВЧАЛЬНОЇ ВИБІРКИ

Розглянуто важливий аспект у сфері машинного навчання — розпізнавання та синтез зображень для ефективного розширення навчальної вибірки. Наведено методи розпізнавання зображень, виокремлено метод згорткових нейронних мереж. Також виявлено одну з основних проблем у процесі використання цих методів, зокрема недостатню навчальну вибірку для навчання мережі. У сучасному машинному навчанні, де доступ до великих обсягів реальних даних часто обмежений, проблема відсутності навчальних прикладів постає як ключова. У статті проаналізовано методи розширення, які дають змогу збільшити обсяг даних завдяки введенню різноманітних перетворень та модифікацій вихідного набору.

Особливу увагу приділено генеративним моделям, таким як Variation Autoencoder (VAE). Детально досліджено їхню здатність синтезувати нові, реалістичні образи та їх вплив на підвищення ефективності для розпізнавання моделей. Розглянуто здатність створювати реалістичні зображення, зазначено важливість використання таких методів у сценаріях із невеликою кількістю доступних даних.

Приклади та результати досліджень підкреслюють практичність застосування запропонованих методів у різних сценаріях машинного навчання. У статті не тільки розкрито сучасні підходи, а й сформульовано перспективи подальших досліджень у цьому важливому напрямі.

**Ключові слова:** нейронні мережі; варіаційний автокодувальник; розпізнавання зображень; генерація зображень; аугментація; навчальна вибірка.

### Вступ

За останні роки все більше набувають популярності такі напрями, як нейронні мережі та машинне навчання. Вони можуть бути використані для різних завдань, зокрема класифікації, регресії, визначення об'єктів у зображеннях, розпізнавання мови, розроблення прогнозів або ухвалення рішень на основі даних. Але для ефективного та якісного навчання нейромережі потрібно якомога більше навчальних даних. Залежно від завдання це може бути різний матеріал, скажімо, нейромережу, яка розпізнає зображення, необхідно навчати саме на таких зображеннях. Але може виникнути ситуація, коли для якісного навчання нейромережі цих зображень буде недостатньо, і тоді головним завданням буде аугментація навчальної вибірки за допомогою генерації зображень.

**Мета статті** — дослідити методи для розпізнавання та генерації зображень для поліпшення точності визначення зображень через аугментацію навчальної вибірки.

**Аналіз останніх досліджень і публікацій.** Сьогодні для розпізнавання зображень існують різні методи, як-от глибоке навчання, зокрема згорткові нейронні мережі (CNN), котрі здатні ефективно впоратися із завданнями класифікації та визначення об'єктів на зображеннях [1; 2]. Крім того, нині популярні методи, засновані на ознаках (наприклад, SIFT, HOG), а також на використанні передньо навчених моделей, зокрема ResNet, Inception та MobileNet. Згорткові нейронні мережі вважаються найбільш доступними в застосуванні для класифікації об'єктів на зображеннях [3].

У праці [4] досліджено та здійснено порівняння різних архітектур згорткових мереж для розпізнавання символів. Було визначено, що обсяг навчальної вибірки відчутно впливає на надійність розпізнавання символів. Для збільшення обсягу навчальної вибірки можна використати Image Data Generator із пакету TensorFlow, який випадково обертає, масштабує, стискає чи витягує зображення. Також цікавим рішенням може бути застосування генеративних нейромереж, які можуть копіювати стилі чи ознаки зображень із навчальної вибірки для створення унікальних зображень.

У статті [5] розглянуто найпопулярніші типи генеративних мереж: варіаційний автоенкодер (VAE), генеративні змагальні мережі (GAN), авторегресивні мережі (NLM). Як наслідок, було з'ясовано: VAE може бути хорошим вибором, якщо важлива можливість варіювання та робота з латентним простором; GAN може бути оптимальним вибором, якщо найважливішою є якість та реалістичність генерованих зображень; авторегресивні мережі можуть бути ефективними, якщо важлива генерація зображень з огляду на структуру та локальні залежності.

### Основна частина

Для розв'язання проблеми навчання нейромережі з недостатньою вибіркою постає завдання її розширити. Варіаційний автоенкодер може слугувати непоганим вибором в досягненні цієї мети, адже

за його допомогою можна згенерувати зразки з контрольованим виявленням ознак. Варіаційний автокодувальник — це тип генеративної моделі, яка базується на ідеї автокодувальника, але додає до неї байєсівський підхід. Основна мета VAE полягає в тому, щоб навчити модель не тільки ефективно кодувати вхідні дані в латентний простір, а й навчити її генерувати нові приклади, подібні до тих, що були в навчальному наборі. VAE працюють на основі двох основних функцій: функції кодування та функції декодування. Функція кодування перетворює вхідні дані в латентний простір, зазвичай поданий як гауссівський розподіл. Після цього функція декодування перетворює точку в цьому латентному просторі назад в оригінальні дані. Розглянемо моделі, з яких сформовано VAE.

**Модель енкодера.** Вхідне зображення  $x$  енкодується в латентний простір за допомогою параметрів гауссівського розподілу:

$$q_{\phi}(z | x) = N(z | \mu_{\phi}(x), \sigma_{\phi}(x)),$$

де  $\mu_{\phi}(x)$  і  $\sigma_{\phi}(x)$  — відповідно середнє та стандартне відхилення, що залежать від вхідного зображення  $x$ .

**Семплювання з латентного простору.** З латентного розподілу  $q_{\phi}(z | x)$  семплюємо, дістаючи конкретний латентний вектор  $z \sim q_{\phi}(z | x)$ .

**Модель декодера.** Латентний вектор  $z$  декодується у реконструйоване зображення  $x'$ :

$$p_{\theta}(x' | z) = N(x' | \mu_{\theta}(z), \sigma_{\theta}(z)),$$

де  $\mu_{\theta}(z)$  і  $\sigma_{\theta}(z)$  — відповідно середнє та стандартне відхилення, які залежать від латентного вектора  $z$ .

**Обчислення нижньої межі (ELBO).** Для визначення ефективності моделі використовується ELBO, що оцінює різницю між точнішим значенням логарифма ймовірності тестового зображення та нижнім обмеженням цієї ймовірності:

$$\log p_{\theta}(x) \geq E_{q_{\phi}(z | x)}[\log p_{\theta}(x | z)] - D_{KL}(q_{\phi}(z | x) || p(z)),$$

де  $D_{KL}$  — дивергенція Кульбака-Лейблера;  $p(z)$  — апіорний розподіл латентних векторів.

**Загальна функція втрат VAE.** Загальна функція втрат VAE об'єднує ELBO з додатковою регуляризацією, забезпечуючи близькість розподілу в латентному просторі до апіорного розподілу:

$$L(\theta, \phi; x) = -E_{q_{\phi}(z | x)}[\log p_{\theta}(x | z)] + D_{KL}(q_{\phi}(z | x) || p(z)),$$

де  $L$  — функція втрат VAE;  $\theta$  і  $\phi$  — параметри моделі відповідно декодера і енкодера.

Розглянуті моделі оптимізуються для того, щоб генерувати нові приклади, максимізуючи ймовірність відповідних даних у латентному просторі. Це дає їм змогу створювати нові, реалістичні приклади, які можна додати в навчальний набір. Для подання результатів було розроблено мережу, яка в змозі згенерувати зображення (рис. 1, рис. 2).



Рис. 1. Набір із навчальної вибірки



Рис. 2. Згенеровані зображення

Для порівняння результатів та перевірки ефективності запропонованого методу, було створено просту згорткову мережу (CNN), яка навчається розпізнавати зображення. Для навчальної вибірки було взято 200 зображень із бібліотеки MNIST. Далі за допомогою VAE цю вибірку було збільшено вдвічі та подано на вхід згорткової мережі (таблиця).

#### Результати навчання нейромереж

Метод	Кількість зображень на вході	Якість розпізнавання	Втрати
CNN	200	Accuracy: 0,8183	loss: 190,7297
CNN+VAE	200+200 (згенеровані)	Accuracy: 0,8636	loss: 150,016

Отже, як впливає з результатів дослідження, нейромережа, котра навчалася на розширеній вибірці, ефективніша приблизно на 5%. Точність розпізнавання підвищено з 81 до 86%, втрати зменшилися з 190 до 150.

#### Висновки

Варіаційний автокодувальник (VAE) став важливим інструментом у сфері глибокого навчання, забезпечуючи здатність генерувати нові дані та реконструювати наявні. Використовуючи байєсівський підхід, VAE визначає нові горизонти в синтезі зображень та моделюванні невизначеностей у даних. Існує

багато застосувань і розширень варіаційних автокодувальників для пристосовування цієї архітектури до різних сфер та поліпшення її продуктивності. Умовний варіаційний автокодувальник (УВАК) (англ. *conditional variational autoencoder*, CVAE) є розширенням класичного варіаційного автокодувальника (VAE), яке дає можливість керувати генерацією даних за певними домовленостями чи контекстом. В основі CVAE лежить ідея використання умови для введення додаткової інформації у процес генерації [6]. Також є архітектури, що поєднують структури варіаційних автокодувальників та породжувальних змагальних мереж (GAN), щоб отримувати гібридні моделі з високою породжувальною здатністю [7].

#### Список використаної літератури

1. Rajavelu A., Musavi M. T., Shirvaikar M. V. *A neural network approach to character recognition // Neural Networks*. 1989. 2(5). P. 387–393. URL: [https://doi.org/10.1016/0893-6080\(89\)90023-3](https://doi.org/10.1016/0893-6080(89)90023-3).
2. *Image character recognition using deep convolutional neural network learned from different languages* / J. Bai, Zh. Chen, B. Feng, Bo. Xu // *IEEE International Conference on Image Processing (ICIP 2014) October 27-30, 2014, Paris, France*. P. 2560–2564.
3. Maitra D. S., Bhattacharya U., Parui S. K. *CNN based common approach to handwritten character recognition of multiple scripts // 3th International Conference on Document Analysis and Recognition (ICDAR)*. August 23-26, 2015, Tunisia, Tuni. P. 1021–1025.
4. *Розпізнавання рукописних українських літер та цифр з використанням синтетичного набору даних та згорткових нейронних мереж* / Є. А. Чичкар'єв, О. В. Зінченко, О. Ю. Балалаєва [та ін.]. 2022. 251 с. URL: <https://doi.org/10.36074/grail-of-science.23.12.2022.36>.
5. Lohvin A. O. *Types of generative neural networks // Scientific notes of Taurida National V. I. Vernadsky University. Series: Technical Sciences*. 2021. Vol. 1, no. 1. P. 102–109. URL: <https://doi.org/10.32838/2663-5941/2021.1-1/17>.
6. Kihyuk S., Honglak L., Xinchun Y. *Learning Structured Output Representation using Deep Conditional Generative Models*. URL: <https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf>.
7. *CVAE-GAN: Fine-Grained Image Generation Through Asymmetric Training* / B. Jianmin, C. Dong, W. Fang [et al.]. 2017. URL: <https://arxiv.org/abs/1703.10155>.

O. Kotul, O. Senkov, O. Zhydko

#### ANALYSIS OF METHODS FOR IMAGE RECOGNITION AND IMAGE SYNTHESIS FOR EXPANDING THE TRAINING SAMPLE

The article examines in detail an important aspect in the field of machine learning — image recognition and synthesis for effective training sample expansion. Methods for image recognition are considered, the method of convolutional neural networks is highlighted. One of the main problems when using these methods was also revealed, namely the insufficient training sample for training the network. In modern machine learning, where access to large amounts of real data is often limited, the problem of lack of training examples emerges as a key one. The importance of data expansion is undeniable in conditions of limited access to a large amount of real data. The article discusses and analyzes augmentation methods that allow you to increase the amount of data by introducing various transformations and modifications to the original set.

Particular emphasis is placed on generative models such as the Variational Autoencoder (VAE). The authors consider in detail their ability to synthesize new, realistic images and their impact on improving the quality of model training. The ability to create realistic and context-sensitive images is examined, considering their pros and cons in the context of expanding training data. The importance of using such methods in scenarios with a small amount of available data is emphasized.

The paper also examines the effect of balanced synthesis and recognition on the performance of models, taking into account the representativeness and diversity of the data. Examples and research results highlight the practicality of using the considered methods in various machine learning scenarios. The purpose of the article is not only to consider existing approaches, but also to point to the prospects and directions of further research in this important direction.

**Keywords:** neural networks; variational autoencoder; image recognition; image generation; augmentation; training sample.