DOI 10.31891/2307-5732-2025-347-8
УДК 004.8

BONDAR VITALII
Cherkasy State Technological University
https://orcid.org/0009-0002-3230-5979
e-mail: v.v.bondar.asp24@chdtu.edu.ua
BABENKO VIRA
Cherkasy State Technological University
https://orcid.org/0000-0003-2039-2841
e-mail: v.babenko@chdtu.edu.ua

INVESTIGATION OF THE MAIN ASPECTS OF THE GENERATIVE MODELS APPLICATIONS IN MACHINE LEARNING

This paper examines the fundamental characteristics and applications of generative deep learning models in machine learning, with particular focus on their methodological frameworks and practical implementations. To identify and analyze the implementation and application characteristics of generative models, we conduct a systematic examination of four main types: autoregressive models, which have become foundational for large language models and sequential data processing; variational autoencoders, which utilize latent space constraints and probabilistic encoding; generative adversarial networks, which employ discriminator-based learning through competitive neural architectures; and diffusion models, which implement a noise-reduction approach through iterative refinement. Each model type presents distinct methodological solutions to the challenge of generating objects that approximate the probability distribution of a given dataset, offering unique advantages and considerations for different application scenarios. This analysis contributes to the understanding of how different generative architectures can be effectively utilized in various machine learning applications.

The research demonstrates the practical applications of these models across multiple domains, highlighting their impact on contemporary machine learning tasks. In computer vision and multimedia, they have proven effective for image synthesis, super-resolution enhancement, and video generation, contributing to significant advances in content creation and modification. Their application extends to healthcare, where they facilitate the generation of synthetic patient data while maintaining privacy requirements and adhering to ethical guidelines. The study highlights the particular utility of generative models in data augmentation scenarios, especially in fields where data collection faces practical or ethical limitations, such as medical imaging and specialized research domains. Special attention is given to the conditioning mechanism that enables natural language interaction with the generation process, which has led to significant advances in text-to-image and text-to-video applications.

Keywords: machine learning, generative models, characteristics, application, data augmentation scenarios.

БОНДАР ВІТАЛІЙ, БАБЕНКО ВІРА
Черкаський державний технологічний університет

ДОСЛІДЖЕННЯ ОСНОВНИХ АСПЕКТІВ ЗАСТОСУВАННЯ ГЕНЕРАТИВНИХ МОДЕЛЕЙ В МАШИННОМУ НАВЧАННІ

У даній роботі досліджуються фундаментальні характеристики та способи застосування генеративних моделей глибокого навчання в машинному навчанні. З метою визначення та аналізу особливостей реалізації та застосування генеративних моделей проаналізовано чотири їх основні типи: авторегресійні моделі, які стали основою для великих мовних моделей; варіаційні автоенкодери, що використовують обмеження латентного простору; генеративні змагальні мережі, які застосовують навчання на основі дискримінатора; та дифузійні моделі, що реалізують підхід зменшення шуму. Кожен тип моделі представляє окремі методологічні рішення для завдання генерації об'єктів, які наближають розподіл ймовірностей заданого набору даних. Проведений аналіз сприяє розумінню того, як різні генеративні архітектури можуть бути ефективно використані в різних застосунках машинного навчання.

Дослідження демонструє практичне застосування цих моделей у різних галузях. У комп'ютерному зорі та мультимедіа вони довели свою ефективність для синтезу зображень, покращення роздільної здатності та генерації відео. Їх застосування поширюється на медичну сферу, де вони сприяють генерації синтетичних даних пацієнтів із дотриманням вимог конфіденційності. У дослідженні підкреслюється особлива корисність генеративних моделей у сценаріях розширення даних, особливо в галузях, де збір даних стикається з практичними чи етичними обмеженнями. Особлива увага приділяється механізму умовної генерації, який забезпечує взаємодію з процесом генерації природною мовою, що призвело до значних досягнень у застосуванні перетворення тексту в зображення та відео.

Ключові слова: машинне навчання, генеративні моделі, характеристики, застосунки, сценарії розширення даних.

Statement of the problem

The rapid advancement of deep learning has led to significant developments in generative models, which have demonstrated the ability to create high-quality synthetic data across various domains, including images, text, music, and other forms of data. While these models show promising results in generating content that closely resembles real-world data, understanding their different approaches and practical applications remains a complex challenge. The growing diversity of generative methodologies, from autoregressive models to diffusion-based approaches, necessitates a systematic examination of their characteristics and capabilities.

This research focuses on analyzing the primary types of generative models and their applications in machine learning tasks. Our investigation addresses the fundamental aspects of these models' architectures, their distinct approaches to data generation, and their practical implementation across different domains. Special attention is given to their role in data augmentation and synthesis, particularly in scenarios where practical or ethical constraints limit data availability.

The purpose of the research is to examine the primary types of generative models in machine learning, analyze their distinct methodological approaches, and explore their practical applications across various domains.

Main material

Generative deep learning models have gained unprecedented popularity in recent years due to their successful application in generating high-quality images, text, music, as well as sensory data, electronic health records, mobility trajectories, and financial time-series [1]. These models have revolutionized various fields by providing innovative solutions for data generation and synthesis. The ability to create realistic and meaningful content has opened new possibilities across multiple domains, from healthcare to creative industries. The growing adoption of generative models has also sparked interest in their practical applications, particularly in scenarios where data augmentation or synthetic data generation is crucial. Along with the successes of generative models in practical applications, a new wave of interest has emerged from researchers, contributing to the continuous appearance of new studies.

Most commonly, a generative model is defined as a model capable of producing new objects that closely resemble those found in a given dataset. In mathematical terms, this process can be understood as approximating the probability distribution of objects within the dataset. In contrast, discriminative models take a fundamentally different approach - instead of working with distributions, they focus on predicting either the likelihood of specific events for individual objects or calculating particular numerical characteristics of these objects [2].

The inherent complexity and ill-defined nature of generative tasks have given rise to numerous and remarkably diverse solution approaches. Among the most widely adopted methodologies in machine learning, we find autoregressive models, variational autoencoders, generative adversarial networks, and diffusion models.

Autoregressive models approach generation by treating the target object as an independent sequence of elements. This fundamental assumption simplifies the task into a more manageable challenge: generating the next element in a sequence based on the existing incomplete subsequence. This approach has proven particularly natural for text generation, where it is commonly known as Language Modeling. Indeed, this methodology became the cornerstone for large language models (LLMs), beginning with the original GPT [3] and continues to be the underlying principle in contemporary models such as Llama [4], GPT-4 [5], Gemini [6] and others.

The variational autoencoder transforms a conventional autoencoder's decoder into a comprehensive generative model by constraining the latent space to conform to a predefined distribution. The original implementation, as presented by Kingma [7], achieved this by constraining the latent space to a unit normal distribution through the ELBO of the log-likelihood. As the field evolved, researchers proposed various alternative approaches to latent space constraint, with discretization-based methods gaining particular prominence, as exemplified by VQ-VAE [8] and FCQ-VAE [9]. However, despite these advances, variational autoencoders continue to face a persistent challenge: the tendency to produce somewhat "blurred" generated objects.

Generative adversarial networks (GANs) replace the formally derived loss function with another neural network, one specifically trained to distinguish between generated objects and actual dataset samples [10]. At the heart of this approach lies the interplay between two networks - the Generator and the Discriminator - which learn to achieve Nash equilibrium through their opposing objectives. While the Discriminator is trained to differentiate between generated and real objects, the Generator is trained to create objects that the Discriminator classifies them as real. In this dynamic relationship, the Discriminator effectively functions as an adaptive loss function for the Generator.

Generative adversarial networks have demonstrated remarkable capabilities across diverse application domains. In computer vision, they have achieved significant success in image-to-image translation, super-resolution enhancement, and the generation of photorealistic images [11]. The medical field has embraced GANs for synthesizing medical images, augmenting limited datasets, and generating synthetic patient data while preserving privacy [12]. In multimedia applications, GANs have proven effective for tasks such as voice conversion, music generation, and video synthesis. The fashion and design industries have utilized GANs for creating new product designs and virtual try-on systems. Additionally, these networks have found applications in anomaly detection, data augmentation for training other machine learning models.

Diffusion generative models frame the generation process as learning to reverse diffusion. In this framework, the forward process entails the gradual addition of white noise to a given object, ultimately transforming it into a point drawn from a normal distribution. The diffusion model itself learns to predict the inverse transformation - the path from a noise-corrupted object back to its original form. Through an iterative process of gradual denoising steps, the model can transform a point sampled from a normal distribution into an object that follows the distribution of the training dataset. While these diffusion processes were traditionally formulated using differential equations [13], recent developments have shown that the same process can be described using simpler mathematical frameworks [14].

Diffusion models gained remarkable popularity following the publication of "Diffusion models beat gans on image synthesis" [15]. These models have since found applications across numerous research areas. In computer vision, researchers have successfully applied them to address fundamental challenges in super-resolution, inpainting, and image restoration. The scope of diffusion models has expanded into video synthesis and point cloud completion tasks. The field of multi-modal learning has also benefited from diffusion models, as evidenced by recent work in text-to-image generation, text-to-video synthesis, and text-to-3D object creation. In the domain of interdisciplinary research, these models have been adapted for molecular graph modeling and have shown practical utility in material design and medical image reconstruction. Their application extends to temporal data analysis, where researchers have employed them for time series imputation, forecasting, and waveform signal processing tasks [16].

In practical tasks where the solution is not uniquely defined (such as classification, regression, and similar problems), generative models prove valuable due to their ability to sample results from a distribution. Various types of generative models are widely employed in the generation of images, video, music, and other media forms. Models that generate content based on textual descriptions or client queries have become particularly prevalent. This conditioning mechanism enables natural interaction with the generation process through natural language expressions. Additionally, generative models have demonstrated considerable utility in enhancement tasks for text, images, video, and audio. Conditioning on lower-quality objects (downscaled, noisy, black-and-white, etc.) provides a strong signal for generative models, and when combined with their capacity to produce high-quality outputs, this becomes the foundation for numerous practical applications.

Another significant aspect of generative model applications lies in the augmentation of small datasets for training specialized neural networks. When faced with limited data availability (for instance, medical images of specific conditions), researchers employ generative models to create realistic synthetic data to supplement their datasets. This approach has proven particularly valuable in domains where data collection is constrained by practical or ethical considerations.

This research examines the application characteristics of four distinct deep learning generative models: autoregressive models, variational autoencoders, generative adversarial networks, and diffusion models. Through a systematic analysis of their fundamental properties and implementation methods, this study provides insights into their practical applications across different fields. The findings from this investigation can assist researchers in selecting appropriate generative models based on specific task requirements within the machine learning framework. This analysis contributes to the understanding of how different generative architectures can be effectively utilized in various machine learning applications.

Conclusions

Generative models have become an integral part of the machine learning landscape. We examined four primary types of generative models: autoregressive models, variational autoencoders, generative adversarial networks, and diffusion models, each offering distinct approaches to the fundamental challenge of generating objects that resemble a given dataset. Autoregressive models have found particular success in text generation and language modeling, serving as the foundation for contemporary large language models. While variational autoencoders have made progress through various approaches to latent space constraint, they continue to face challenges with output clarity. Generative adversarial networks have shown remarkable versatility, while diffusion models have demonstrated promising results across multiple domains.

The practical applications of these models have shown significant impact across various fields. In computer vision and multimedia, they excel at tasks such as image-to-image translation, super-resolution enhancement, and video synthesis. The medical field has benefited from their ability to generate synthetic patient data while preserving privacy. Their natural language conditioning mechanism has enabled intuitive interaction with generation processes, leading to advances in text-to-image generation and text-to-video synthesis. Particularly noteworthy is their role in data augmentation, where they address the challenge of limited data availability by generating realistic synthetic data for training specialized neural networks. This capability has proven especially valuable in domains where data collection is constrained by practical or ethical considerations, such as medical imaging and specialized research fields.

References

1. Eigenschink, P., Reutterer, T., Vamosi, S., Vamosi, R., Sun, C., & Kalcher, K. (2023). Deep generative models for synthetic data: A survey. *IEEE Access, 11*, 47304–47320.

2. Harshvardhan, G. M., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review, 38*, Article 100285.

3. Radford, A. (2018). Improving language understanding by generative pre-training. *arXiv preprint arXiv:1810.04805*.

4. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

5. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). GPT-4 technical report. *arXiv preprint arXiv:2303.08774*.

6. Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., ... & Blanco, L. (2023). Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

7. Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*.

8. Van den Oord, A., Vinyals, O., & Kavukcuoglu, K. (2017). Neural discrete representation learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*.

9. Mentzer, F., Blumer, T., Hoogeboom, E., & Menkovski, V. (2023). Finite scalar quantization: VQ-VAE made simple. *arXiv preprint arXiv:2309.15505*.

10. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144.

11. Wang, Z., She, Q., & Ward, T. E. (2021). Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR), 54*(2), 1–38.

12. Chen, Y., Yang, X. H., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., ... & Guan, Q. (2022). Generative adversarial networks in medical image augmentation: A review. *Computers in Biology and Medicine, 144*, Article 105382.

13. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems, 33*, 6840–6851.

14. Heitz, E., Belcour, L., & Chambon, T. (2023). Iterative α-(de)blending: A minimalist deterministic diffusion model. *ACM SIGGRAPH 2023 Conference Proceedings*.

15. Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems, 34*, 8780–8794.

16. Yang, L., Cheng, X., Zhang, Z., Wang, Y., Luo, H., & Wang, H. (2023). Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys, 56*(4), 1–39.