UDK: 004.8:004.932

Chekhmestruk Roman Y.
Kudran Olha P.

MECHANISM FOR BUILDING DIFFUSION MODELS FOR IMAGE GENERATION

Vinnytsia National Technical University

Анотація

У статті розглядається механізм побудови дифузійних моделей для генерації зображень як один із найсучасніших напрямків штучного інтелекту та комп'ютерного зору. Описано принцип роботи процесів дифузії, який полягає у поступовому зменшенні шуму зображення та його подальшому відновленні за допомогою нейронної мережі. Проаналізовано основні етапи навчання моделі, включаючи процеси прямої та зворотної дифузії, а також особливості архітектури мережі. Розглянуто переваги дифузійних моделей у порівнянні з іншими генеративними підходами, зокрема GAN та VAE, з точки зору якості генерованих зображень та стабільності навчання. Окреслено практичні сфери застосування таких моделей у дизайні, медицині, розвагах та автоматизованому створенні контенту.

Ключові слова: дифузійні моделі, генерація зображень, штучний інтелект, нейронні мережі, комп'ютерний зір, генеративні моделі, процес дифузії, глибоке навчання, синтез зображень, обробка зображень.

Abstract

The paper considers the mechanism of building diffusion models for image generation as one of the most modern areas of artificial intelligence and computer vision. The principle of operation of diffusion processes is described, which consists in the gradual noise reduction of the image and its subsequent restoration using a neural network. The main stages of model training are analyzed, including forward and backward diffusion processes, as well as the features of the network architecture. The advantages of diffusion models in comparison with other generative approaches, in particular GAN and VAE, are considered in terms of the quality of generated images and the stability of training. The practical areas of application of such models in design, medicine, entertainment, and automated content creation are outlined.

Keywords: diffusion models, image generation, artificial intelligence, neural networks, computer vision, generative models, diffusion process, deep learning, image synthesis, image processing.

The mechanism for constructing diffusion models for image generation is considered one of the most significant developments in modern artificial intelligence and computer vision. Diffusion models are based on probabilistic principles that allow neural networks to learn how to transform random noise into meaningful visual structures. During training, the model gradually learns the statistical distribution of images contained in a dataset and uses this information to synthesize new images with similar properties. This approach has become particularly important in recent years due to its ability to produce highly realistic and diverse images while maintaining stable training processes compared to earlier generative methods [1].

The fundamental concept of diffusion models is the forward diffusion process, in which noise is progressively added to an image through a sequence of steps until the original signal is almost completely destroyed. Each step slightly modifies the data distribution, eventually converting the image into nearly pure random noise. This controlled degradation process forms the basis for learning how images can be reconstructed during the reverse diffusion stage [2].

The reverse diffusion process performs the opposite operation by gradually removing noise from a corrupted image. A neural network is trained to predict and eliminate the noise introduced during the forward process. Through multiple denoising steps, the model reconstructs structured visual information from random noise, effectively generating new images that resemble the distribution of the training data [3].

Recent studies demonstrate that diffusion models can achieve impressive performance in various domains, including medical imaging and digital modeling. In particular, they can be used to modify or generate complex visual structures while preserving anatomical or structural consistency, which opens new possibilities for applications in scientific research and medical analysis [4].

Another important advantage of diffusion models is their flexibility in practical applications. These models can be applied in many fields such as digital design, architecture, simulation, and content creation. Due to their ability to produce detailed and coherent visual outputs, diffusion models are increasingly used in modern visual computing systems and creative industries [5].

To improve computational efficiency, modern systems often implement latent diffusion models. In this approach, the diffusion process operates not directly on high-resolution pixel data but on compressed latent representations of images. This significantly reduces the computational resources required for training and inference while still preserving the essential visual features of the generated images [6].

From an architectural perspective, diffusion models commonly rely on convolutional neural networks such as U-Net. This architecture combines encoder and decoder structures with skip connections that preserve spatial information across multiple resolution levels. Such design enables the model to effectively capture both global image structures and fine visual details, which is essential for generating high-quality images [7].

Conclusion

In this study, the fundamental principles of diffusion models for image generation were analyzed. The mechanism of their operation is based on a stochastic process that gradually adds noise to data and then restores the original structure through a neural network. Such an approach enables the generation of realistic images that correspond to the statistical distribution of the training dataset.

The analysis showed that diffusion models provide several advantages compared to other generative approaches, including high stability during training, the ability to reproduce complex visual structures, and the generation of diverse and detailed images. The use of latent representations and neural network architectures such as U-Net further improves the efficiency and quality of image synthesis.

Overall, diffusion models represent a promising direction in the development of generative artificial intelligence systems. Their continuous improvement and integration with other modern machine learning techniques create new opportunities for automated visual content generation and various practical applications in science, technology, and creative industries.

LIST OF REFERENCES

1. Google Research. Imagen: Text-to-Image Diffusion Models. 2022. Тема: принцип роботи diffusion-моделей для генерації зображень за текстовим описом. С. 1–5.
2. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. 2020. Тема: математична модель diffusion-процесу та алгоритм денойзингу у генеративних моделях. С. 6840–6851.
3. IBM. What Are Diffusion Models? 2023. Тема: загальний принцип роботи diffusion-моделей та сфери їх застосування у штучному інтелекті. С. 1–4.
4. Kadry K., Thouvenin P.-A., Salvador T., Nadeau M., Kroll J. D., Tam A., Ostmeier S., Beaulieu-Jones B. K., Shen C., Yetisir E., Acosta J. N., Bertsimas D., Vazquez M., Mouchess M., Kalpathy-Cramer J. Probing the Limits and Capabilities of Diffusion Models for the Anatomic Editing of Digital Twins. 2024. Тема: використання diffusion-моделей для редагування медичних зображень та цифрових двійників. С. 1–12.
5. NVIDIA Developer. Understanding Diffusion Models: An Essential Guide for AEC Professionals. 2023. Тема: архітектура diffusion-моделей та їх використання у генерації зображень. С. 1–6.
6. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models. 2022. Тема: latent diffusion models та генерація високоякісних зображень. С. 10674–10685.
7. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. Тема: архітектура U-Net для сегментації медичних зображень у задачах комп’ютерного зору. С. 234–241.

Чехместрук Роман Юрійович – доцент кафедри програмної інженерії, Вінницький національний технічний університет, Вінниця, Україна, e-mail: chekhroma@gmail.com

Кудран Ольга Петрівна – студентка факультету ІТКІ, Вінницький національний технічний університет, Вінниця, e-mail: olha.kudran@gmail.com

Chekhmestruk Roman Y. – Associate Professor of the Department of Software Engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine, e-mail: chekhroma@gmail.com

Kudran Olha P. – student of the Faculty of ITKI, Vinnytsia National Technical University, Vinnytsia, e-mail: olha.kudran@gmail.com