

# Generating images using generative adversarial networks based on text descriptions

Marzhan Turarova<sup>1</sup>, Roza Bekbayeva<sup>2</sup>, Lazzat Abdykerimova<sup>3</sup>, Murat Aitimov<sup>4</sup>, Aigulim Bayegizova<sup>5</sup>, Ulmeken Smailova<sup>6</sup>, Leila Kassenova<sup>7</sup>, Natalya Glazyrina<sup>1</sup>

<sup>1</sup>Department of Computer and Software Engineering, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan

<sup>2</sup>Department of Automation, Information Technology and Urban Development of Non-Profit Limited Company Semey University named after Shakarim, Semey, Republic of Kazakhstan

<sup>3</sup>Department of Information Systems, M.H. Dulaty Taraz Regional University, Taraz, Republic of Kazakhstan

<sup>4</sup>Kyzylorda Regional Branch at the Academy of Public Administration under the President of the Republic of Kazakhstan, Kyzylorda, Republic of Kazakhstan

<sup>5</sup>Department of Radio Engineering, Electronics and Telecommunications, L. N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan

<sup>6</sup>Center of Excellence of Autonomous Educational Organization Nazarbayev Intellectual Schools, Astana, Republic of Kazakhstan

<sup>7</sup>Department of Information Systems and Technologies, Faculty of Applied Sciences, Esil University, Astana, Republic of Kazakhstan

## Article Info

### Article history:

Received Nov 1, 2023

Revised Dec 12, 2023

Accepted Jan 5, 2024

### Keywords:

Discriminator

Extra-long transformer network

Generative adversarial network

with conditional latent semantic

Generator

Machine learning

Natural language processing

## ABSTRACT

Modern developments in the fields of natural language processing (NLP) and computer vision (CV) emphasize the increasing importance of generating images from text descriptions. The presented article analyzes and compares two key methods in this area: generative adversarial network with conditional latent semantic analysis (GAN-CLS) and ultra-long transformer network (XLNet). The main components of GAN-CLS, including the generator, discriminator, and text encoder, are discussed in the context of their functional tasks—generating images from text inputs, assessing the realism of generated images, and converting text descriptions into latent spaces, respectively. A detailed comparative analysis of the performance of GAN-CLS and XLNet, the latter of which is widely used in the organic light-emitting diode (OEL) field, is carried out. The purpose of the study is to determine the effectiveness of each method in different scenarios and then provide valuable recommendations for selecting the best method for generating images from text descriptions, taking into account specific tasks and resources. Ultimately, our paper aims to be a valuable research resource by providing scientific guidance for NLP and CV experts.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Roza Bekbayeva

Department of Automation, Information Technology and Urban Development of Non-Profit Limited

Company Semey University named after Shakarim

Semey, Republic of Kazakhstan

Email: rbekbayeva@internet.ru

## 1. INTRODUCTION

Integrating text descriptions with visual data is one of the most current and important tasks in the field of artificial intelligence (AI) [1]–[3] and computer vision (CV) [4]–[6]. The ability to generate images from textual descriptions [7] has the potential to change the ways we perceive and interact with the world of visual data. In this paper, we explore two leading methods designed to solve this problem: generative adversarial network with conditional latent semantic analysis (GAN-CLS) [8] and extra-long transformer network (XLNet)

[9]. The relevance of this topic is difficult to overestimate. With the advent of large volumes of text data and the availability of powerful computing resources, there has been an increased need to develop algorithms that can analyze and interpret text descriptions and convert them into visual images. This has huge application potential in various fields such as medical diagnostics [10], automatic art generation [11], education, entertainment and many others. The purpose of this article is to conduct an in-depth study and comparative analysis of two methods that are of interest for solving the problem of generating images based on text descriptions. We strive to identify their advantages, disadvantages and applications.

The GAN-CLS method [12], [13] is a combination of a generative adversarial network (GAN) and a text encoding module. The main idea is to use conditional space to improve the quality of image generation based on text descriptions. The generator of this architecture is built on a sequence of convolutional transposed layers, and the discriminator serves to classify images into real and generated. The text encoder [14] plays an important role in providing the connection between text and visual representation. XLNet, on the other hand, is a transformer model that uses an attention mechanism to encode input data. It differs from standard autoregressive models such as bidirectional encoder representations from transformers (BERT) in that it predicts each token based on all other tokens in the sequence, not just the previous ones. This allows you to take into account more complex dependencies in the data and better analyze the text. In our research, we conduct a detailed analysis of the architecture and performance of both methods based on experiments and expert opinion. We identify in which scenarios and for which tasks each method may be most useful, and identify the potential limitations and challenges they may face.

Berrahal and Azizi [15] describes a research project focused on applications of image-from-text synthesis, especially in the area of generating images of people's faces from text descriptions. The researchers used unsupervised deep neural networks, with a focus on deep fusion generative adversarial networks (DF-GANs), which outperformed other approaches in the quality of generated images and correspondence to text descriptions. The main goal of this research appears to be to assist law enforcement agencies by creating realistic and diverse facial images based on eyewitness descriptions. The article emphasizes that the implemented model demonstrates excellent quantitative and visual characteristics, and is capable of creating realistic and diverse facial images while respecting the entered text descriptions. This technology could potentially be valuable in creating profiles for law enforcement agencies based on eyewitness testimony.

Luo *et al.* [16] addresses the problem of synthesizing facial images from text descriptions. The authors present a new method, dual-channel generator based generative adversarial network (DualG-GAN), based on generative adversarial networks. The main goal is to improve the quality of the generated images and their consistency with the text description. A two-channel generator is introduced to improve consistency, and a new loss function is developed to improve the similarity between the generated image and the real one at three semantic levels. Experiments show that DualG-GAN achieves the best results on the text-to-face (SCU-Text2face) dataset and outperforms other methods in text-to-image synthesis tasks. Ku and Lee [17] presents a TextControlGAN model based on GANs specifically designed for synthesizing images from text descriptions. Unlike conventional GANs, TextControlGAN includes a neural network structure, a regressor, to efficiently extract features from conditional texts. The model also uses data augmentation techniques to improve regressor training. This allows the generator to more efficiently learn conditional texts and create images that more closely match the text descriptions.

Cheng *et al.* [18] addresses the problem of image synthesis from text descriptions, proposing a novel vision-language matching GAN (VLMGAN) strategy to improve the quality and semantic consistency of the generated images. The model, called VLMGAN, introduces a dual mapping mechanism between text and image to strengthen image quality and semantic consistency. This mechanism takes into account the mapping between the generated image and the text description, as well as the mapping between the synthesized image and the real image. The proposed strategy can be applied to other text-to-image synthesis methods. Chopra *et al.* [19] provide an overview of current advances in the field of image synthesis from text descriptions. The authors study various architectures of generative models designed to solve the problem of image synthesis from a text description. There has been significant progress in the field of artificial intelligence, especially in the field of deep learning, which has led to the creation of generative models capable of creating realistic images based on text descriptions. The article reviews standard GANs, deep convolutional GAN (DCGAN), stacked GAN (StackGAN), StackGAN++, and attention GAN (AttnGAN). Each of these models presents different approaches and improvements for synthesizing images from text, including the use of attention and iterative refinement.

## 2. METHOD

Generating images from text descriptions [20]–[22] is a complex and multifaceted task that has attracted the attention of researchers in the fields of artificial intelligence and computer vision for many years. In this section, we review some of the key methods proposed in the literature and analyze their main

characteristics. One of the early and important approaches to generating images from text descriptions was a method based on conditional GANs (cGANs) [23]–[25]. This method involves the use of GANs with conditional input, which is a text description. Our research includes a GAN architecture specialized for working with text information and images. This architecture is based on the process of encoding textual information, which we denote as  $\phi(t)$ . This encoding process is performed on both the generator, which is responsible for creating the images, and the discriminator, which determines how realistic the created images are.

The main characteristic of this architecture is the integration of textual information with visual features of images. To do this, the text encoding  $\phi(t)$  is projected to a lower dimension and depth to reduce dimensionality and improve data processing. The resulting textual representation is then concatenated with image feature maps to combine textual and visual information. Next, both the generator as shown in Figure 1(a) and the discriminator as shown in Figure 1(b) operate on this combined representation. The generator uses it to create images that match text descriptions, making the process more controlled and conditional. The discriminator uses this representation to evaluate the generated images and determine how realistic they are. Thus, this architecture provides interaction between text and visual information, which makes it a powerful tool for the task of generating images based on text descriptions.

The generator in cGAN learns to convert this text input into an image, and the discriminator tries to distinguish between real and generated images. This method has several advantages, but also faces challenges related to text and image matching. Another popular approach is to use Autoencoders to generate images from text. In this case, the autoencoder is trained to transform a textual description into a latent representation, and then reconstruct an image from this representation. This method also has its strengths, but may encounter problems due to the limited ability of the model to produce varied and high-quality images. With the advent of transformer models such as BERT and generative pre-trained transformer (GPT), the field of natural language processing (NLP) and computer vision has a new opportunity to generate images from text. These models have shown impressive results in tasks related to text analysis and have become a source of inspiration for researchers in the field of image generation. In particular, the XLNet method, which we discuss in this paper, uses a transformer architecture and an attention mechanism for text analysis and image generation. However, despite significant progress in this field, the task of generating images from text descriptions remains a challenge. Questions arise related to the quality of the generated images, the variety of content created, the interpretation of text descriptions and other aspects. This makes it one of the current research areas in the field of artificial intelligence and computer vision. In the following sections of our article, we will take a closer look at the GAN-CLS and XLNet method, conduct experiments to compare their performance, and discuss the results of the study.

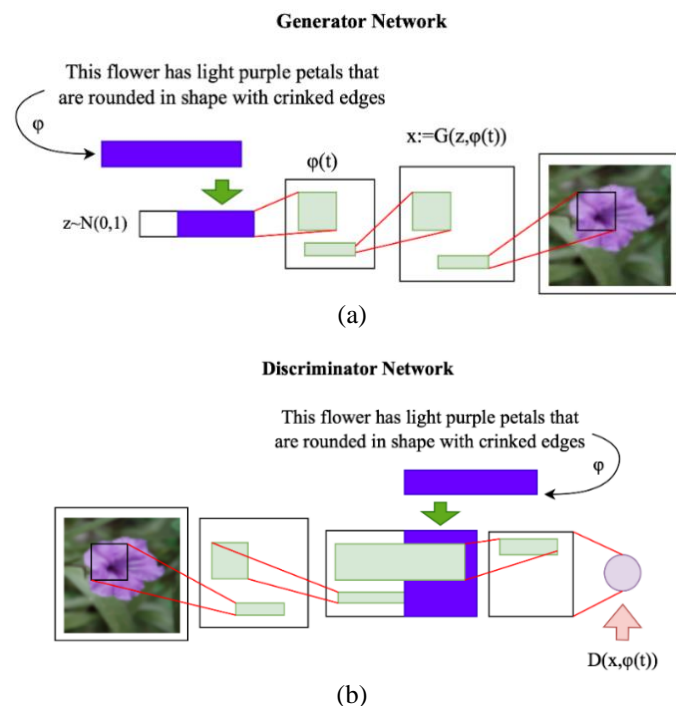


Figure 1. Architecture of cGAN (a) generator and (b) discriminator

### 3. RESULTS AND DISCUSSION

In this research work, a comparative analysis of two methods was carried out: GAN-CLS and XLNet, used in the task of generating images based on text descriptions. For this purpose, a dataset containing textual descriptions and corresponding images was used, which was divided into training and testing sets. Before training the models, text descriptions were preprocessed, including tokenization and conversion of words into numeric representations. GAN-CLS is based on a GAN architecture integrated with a text encoding module. The work analyzed the main components of GAN-CLS, including the generator, discriminator and text encoder. The results showed that the generator loss ranged from 0.1273 to 0.9893, with a mean value of about 0.4807, while the discriminator loss remained stable with a mean value of about 1.4696 as shown in Figure 2.

XLNet, with its transformer architecture and attention mechanism, has shown an impressive ability to quickly reduce generator losses. The average value of about 5.8187 indicates the effectiveness of this process, despite the fact that the absolute values of losses were higher. It is important to note that even with this increase in absolute values, the rate of decrease in losses indicates the high learning ability of the model. On the other hand, discriminator losses in XLNet decreased less uniformly, which may indicate more difficult challenges in training the discriminative part of the model. An average value of about 0.7353 as shown in Figure 3 indicates a general decreasing trend in losses, but additional fine tuning may be required to achieve a more stable and uniform reduction in discriminator losses during training.

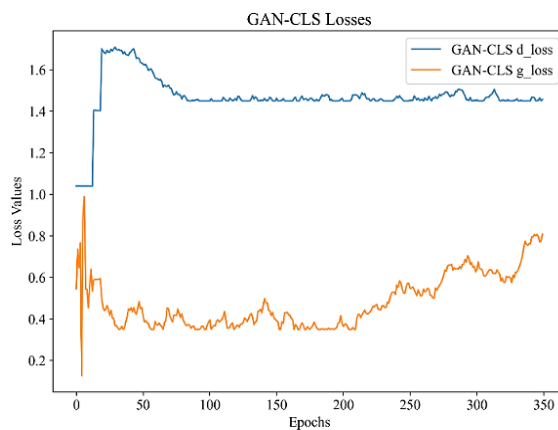


Figure 2. GAN-CLS loss plot

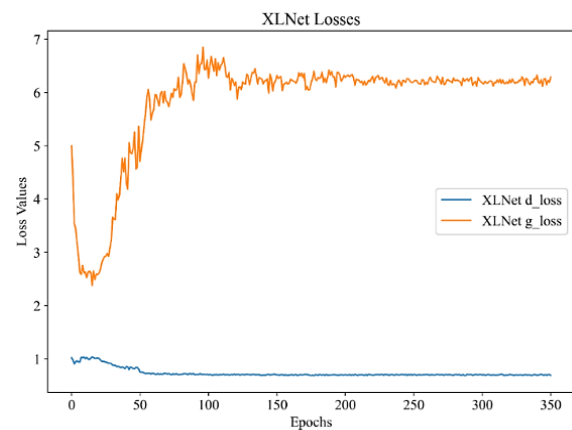


Figure 3. Loss graph of a GAN model with an XLNet encoder

Comparing the performance of both methods, GAN-CLS demonstrated stable training with low generator loss, indicating its ability to generate high-quality images from text descriptions. While XLNet, due to its transformer architecture and attention mechanism, was able to capture both short-term and long-term dependencies in the data, although the absolute loss values were higher. Architectural complexity also plays an important role: GAN-CLS is a simpler model suitable for resource-constrained scenarios and Text2Image tasks, while XLNet, although capable of providing deeper text analysis and accurate results, is computationally and time-consuming for training. Thus, the choice between these methods depends on the specific task and available resources, researchers and practitioners should consider these factors when choosing a method for generating images from text descriptions.

Continuing our research in the field of image generation from text descriptions, we focus on a more detailed analysis of the GAN-CLS and XLNet methods, considering their advantages and limitations in various aspects. GAN-CLS is an innovative approach that successfully integrates generative adversarial networks with a text encoding module. One of the main advantages of GAN-CLS is its ability to generate high-quality images based on text descriptions. This is made possible by using a conditional space for latent semantic analysis, which allows the model to take into account the semantic information of the text when generating images. The generator in GAN-CLS consists of a sequence of convolutional transpose layers, which allows the image resolution to increase as it is generated. These rectified linear unit (ReLU) activation layers help create complex and detailed patterns in images. The final generator layer uses the Tanh activation function to normalize the output values. The discriminator in GAN-CLS is used to classify images as real or generated. This is achieved through a sequence of convolutional layers, including batch normalization and ReLU activation functions. After convolution, the image is passed through an averaging layer and combined

with text information for classification. The text encoder plays a key role in the relationship between the text description and the image. It uses an embedding layer to transform words into vector representations and a bidirectional GRU to parse text sequences.

The advantages of GAN-CLS include stable training, specialization for the Text2Image task, and low generator loss. However, this method may not capture more complex semantic features of text compared to more complex models such as XLNet. On the other hand, it is a powerful transformer model that uses the attention mechanism to encode input data. The main advantage of XLNet is its ability to take into account information from all parts of the input data when predicting each token in the sequence, making it capable of capturing both short-term and long-term dependencies in the data. XLNet consists of multiple layers of attention, which allows it to capture complex dependencies in data. Unlike some other models, such as BERT, XLNet takes into account information from all parts of the input data when predicting each token. Benefits of XLNet include rapid reduction of generator losses and the ability to better capture complex dependencies in data. However, it has high absolute values of generator losses and requires more computational resources and training time. To summarize, the choice between GAN-CLS and XLNet depends on the specific task and available resources. GAN-CLS provides a simple and efficient way for Text2Image tasks, while XLNet provides in-depth text analysis and accurate results. It is important to consider both the quality of the generated images and the resources required to train and apply the techniques when choosing the best approach for generating images from text descriptions.

The GAN-CLS architecture starts with a generator that consists of a sequence of convolutional transpose layers. These layers incrementally increase the image resolution using batch normalization and ReLU activation functions to introduce nonlinearity. The final generator layer uses the Tanh activation function to normalize values between -1 and 1, which is standard for normalized images. The discriminator, on the other hand, classifies the images as real or generated using convolutional layers and batch normalization, and after convolution, the images are passed through an averaging layer. The text encoder plays a key role in this architecture, using an embedding layer to transform word indices into vector representations and a bidirectional GRU to parse the text sequence. Thus, GAN-CLS integrates a GAN with a text encoding module to improve the quality of image generation based on text descriptions. On the other hand, XLNet is a transformer model that uses an attention mechanism to encode input data. Its architecture includes multiple layers of attention to capture complex dependencies in data. Bidirectional attention is one of the key features of XLNet, allowing the model to consider information from all parts of the input data when predicting each token. The transformer structure on which XLNet is based is much more complex and powerful than the RNN architecture. Multiple layers of attention and parallel data processing make XLNet an effective tool for analyzing text and modeling complex dependencies in data. Both architectures provide powerful tools for working with text and generating images from text descriptions, but the choice between them depends on specific tasks and resource requirements. In this study, an extensive dataset as shown in Figure 4, which consisted of images and text descriptions, was used to train and evaluate the performance of GAN-CLS and XLNet methods. This dataset included pairs of images, where each pair contained one correct image and one incorrect image. These pairs served as the basis for training and evaluating the performance of the models. For each correct image in the dataset, a text description was provided that contained a detailed description of the content of the image. This description played a crucial role in the image generation task, since the models had to learn to create images based on text descriptions. The correct images provided visual context consistent with the text descriptions. The main focus was on correctly annotating the data in the dataset to ensure that the models received high-quality training information and were able to identify correspondence between text descriptions and images. This dataset played a key role in training and testing the GAN-CLS and XLNet methods, and allowed for a comparative analysis of their performance in the context of the problem of generating images based on text descriptions.

The Oxford 102 Flower database is a collection of the 102 flower categories found most commonly in the UK. This database is used for image classification tasks and includes a variety of color categories, each containing from 40 to 258 images. However, there are several important aspects to consider when analyzing this database. First, the images in this database are subject to various variations, such as different scale, viewing angle, and lighting. This makes the classification task more complex and requires image processing algorithms to be robust to such changes. In addition, some color categories in this database may have significant within-category variations. This means that within the same flower category there may be images with different characteristics, such as the color palette, shape and structure of the flower. This further complicates the classification task and requires more accurate and versatile image analysis methods. It should also be noted that there are several very similar color categories in the database as shown in Figure 5. This means that some flowers may be visually similar and distinguishing them may be difficult even for humans. This presents an additional challenge for machine learning algorithms because they have to separate similar classes.

The images presented are the result of the combined use of a GAN algorithm and an autoencoder, which uses the XLNet model inside. This approach has demonstrated outstanding results in interpreting input textual descriptions, demonstrating its effectiveness and accuracy in matching textual data with visual elements of images. There are several key factors to note that make this method more attractive and best in class. First, using GAN allows you to create images that visually match descriptions by training a generator on input text data and visual data, thus achieving high consistency between text and image. Secondly, the autoencoder, which includes the XLNet model, provides a more accurate and in-depth understanding of the entered words and their relationships. XLNet is a high-performance natural language architecture that helps you analyze text data more accurately and better match visual elements in images. Thus, the combination of GAN and autoencoder using the XLNet model is an advanced method that not only provides outstanding accuracy and connectivity between textual and visual data, but also promotes a deeper and more accurate understanding of the context of the input words. This method is best in class, achieving impressive results in interpreting text descriptions and visualizing them in images as shown in Figure 6.

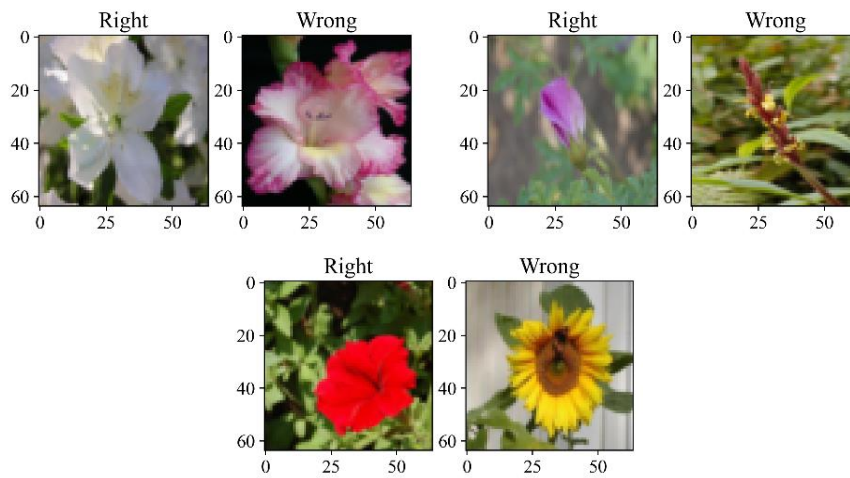


Figure 4. Training dataset

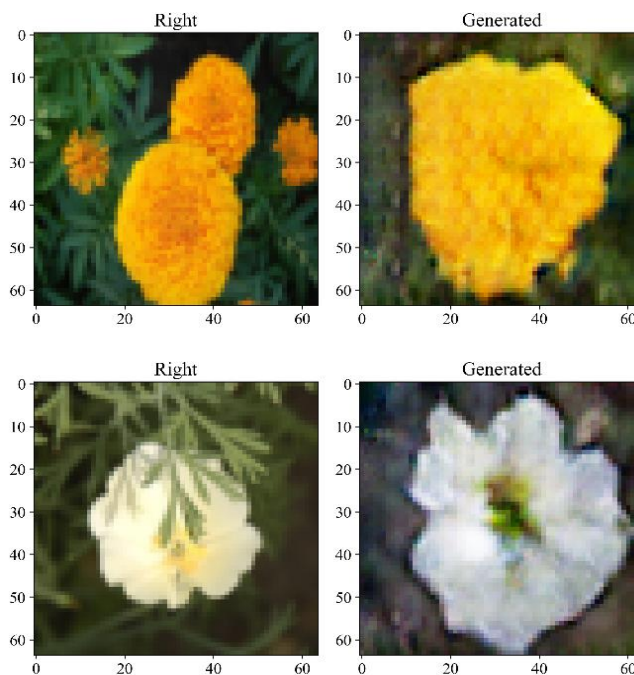


Figure 5. Similar color categories in the database

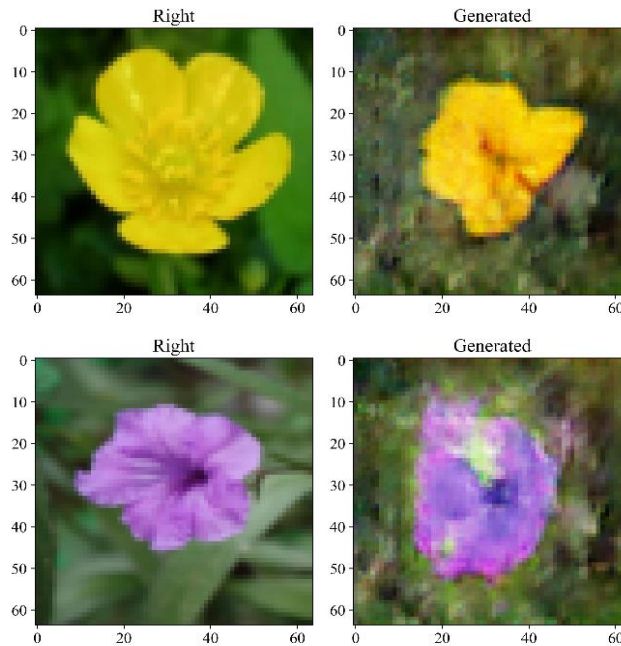


Figure 6. Results in the interpretation of text descriptions and their visualization in images

The presented images show the results of the GAN-CLS algorithm, which uses an autoencoder based on gated recurrent unit (GRU). Obviously, this method performed less effectively compared to the previous version based on GAN-XLNet. There are several factors that explain why this method was less successful. First, an autoencoder built on GRU may not have the same ability to understand and interpret text data as XLNet. XLNet is a more advanced architecture for working with natural languages and is better able to capture the semantic relationships and context of text descriptions, which is important when creating correspondence between text and images. Second, it is possible that the GRU-based autoencoder does not have the same ability to learn more complex shapes and structures of objects in images. This can result in a less accurate recreation of the shape of objects, as can be seen in the provided images where the shape does not match expectations. Thus, the relatively less successful results of GAN-CLS using the GRU autoencoder can be explained by the model's limited ability to analyze text data and learn complex visual features. In this context, the GAN-XLNet algorithm, which uses XLNet and GAN, is a more powerful and efficient method for generating correspondence between text and image with high accuracy and quality.

#### 4. CONCLUSION

In conclusion of this extensive research work, we would like to summarize and summarize the main results, as well as discuss current prospects and directions for future research in the field of image generation based on text descriptions, using two methods: GAN-CLS and XLNet. The purpose of this article was to conduct a comparative analysis of these two methods and determine their advantages and disadvantages in the context of an image generation problem. We started with a detailed look at the GAN-CLS architecture, which integrates GANs with a text encoding module. We reviewed the key components of this architecture, including the generator, discriminator, and text encoder, and analyzed its performance. We found that GAN-CLS exhibits stable learning and specification for the Text2Image task, but may be limited in capturing more complex semantic features of text. We then moved on to the transformer model-based XLNet architecture and analyzed its ability to capture complex dependencies in text. We find that XLNet has a complex architecture with multiple attention layers and bidirectional attention, which allows it to efficiently analyze text and take into account context from the entire sequence. After comparing the two methods, we found that the choice between them depends on specific tasks and resources. GAN-CLS may be preferable in cases where stable training and specification for Text2Image is required, but it may be limited in the complexity of text analysis. On the other hand, XLNet can offer deeper text analysis and capture complex dependencies more efficiently, but it requires more computational resources and training time.

So, summarizing our results, we emphasize that each method has its own strengths and weaknesses. The choice of method depends on the specific task, resources and required accuracy. We hope that this paper

will provide important guidance for optimal method selection in the field of image generation from text descriptions and stimulate further research in this exciting area to create more efficient and innovative methods. Thus, research in the field of image generation based on text descriptions remains relevant and holds great promise, especially given the rapid development of machine learning and artificial intelligence methods. The development of more complex and efficient models capable of capturing deeper semantic dependencies in text is one of the key directions for future research in this area. We expect that the collaborative efforts of researchers and engineers will lead to new advances and innovative applications in image generation from text descriptions.




## REFERENCES

- [1] A. El-Komy, O. R. Shahin, R. M. Abd El-Aziz, and A. I. Taloba, "Integration of computer vision and natural language processing in multimedia robotics application," *Information Sciences Letters*, vol. 11, no. 3, pp. 765–775, May 2022, doi: 10.18576/isl/110309.
- [2] Y. Bian, Y. Lu, and J. Li, "Research on an artificial intelligence-based professional ability evaluation system from the perspective of industry-education integration," *Scientific Programming*, vol. 2022, pp. 1–20, Aug. 2022, doi: 10.1155/2022/4478115.
- [3] T. H. Lin, Y. H. Huang, and A. Putranto, "Intelligent question and answer system for building information modeling and artificial intelligence of things based on the bidirectional encoder representations from transformers model," *Automation in Construction*, vol. 142, Oct. 2022, doi: 10.1016/j.autcon.2022.104483.
- [4] M. H. Guo *et al.*, "Attention mechanisms in computer vision: A survey," *Computational Visual Media*, vol. 8, no. 3, pp. 331–368, Mar. 2022, doi: 10.1007/s41095-022-0271-y.
- [5] L. Zhou, L. Zhang, and N. Konz, "Computer vision techniques in manufacturing," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 105–117, Jan. 2023, doi: 10.1109/TSMC.2022.3166397.
- [6] I. Gao, G. Ilharco, S. Lundberg, and M. T. Ribeiro, "Adaptive testing of computer vision models," *arXiv:2212.02774*, 2022.
- [7] A. Bayegizova *et al.*, "Effectiveness of the use of algorithms and methods of artificial technologies for sign language recognition for people with disabilities," *Eastern-European Journal of Enterprise Technologies*, vol. 4, no. 2–118, pp. 25–31, Aug. 2022, doi: 10.15587/1729-4061.2022.262509.
- [8] J. Li, T. Sun, Z. Yang, and Z. Yuan, "Methods and datasets of text to image synthesis based on generative adversarial network," in *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Sep. 2022, pp. 843–847, doi: 10.1109/ICISCAE55891.2022.9927634.
- [9] N. Habbat, H. Anoun, and L. Hassouni, "Combination of GRU and CNN deep learning models for sentiment analysis on French customer reviews using XLNet model," *IEEE Engineering Management Review*, vol. 51, no. 1, pp. 41–51, Mar. 2023, doi: 10.1109/EMR.2022.3208818.
- [10] G. Abdikerimova *et al.*, "Detection of chest pathologies using autocorrelation functions," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, pp. 4526–4534, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4526-4534.
- [11] Y. Shi, D. Deb, and A. K. Jain, "WarpGAN: automatic caricature generation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 10754–10763, doi: 10.1109/CVPR.2019.01102.
- [12] J. Liu, L. Zheng, X. Zhang, and Z. Guo, "Power grid fault diagnosis method based on alarm information and PMU fusion," in *Fifth International Conference on Computer Information Science and Artificial Intelligence (CISAI 2022)*, Mar. 2023, doi: 10.1117/12.2668200.
- [13] J. Wang, "A text image generation model based on deep learning," *Journal of Intelligent and Fuzzy Systems*, vol. 45, no. 3, pp. 4979–4989, Aug. 2023, doi: 10.3233/JIFS-223741.
- [14] F. Wu *et al.*, "Wav2Seq: Pre-training speech-to-text encoder-decoder models using pseudo languages," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10096988.
- [15] M. Berrahal and M. Azizi, "Optimal text-to-image synthesis model for generating portrait images using generative adversarial network techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 25, no. 2, pp. 972–979, Feb. 2022, doi: 10.11591/ijeecs.v25.i2.pp972-979.
- [16] X. Luo, X. He, X. Chen, L. Qing, and J. Zhang, "DualG-GAN, a dual-channel generator based generative adversarial network for text-to-face synthesis," *Neural Networks*, vol. 155, pp. 155–167, Nov. 2022, doi: 10.1016/j.neunet.2022.08.016.
- [17] H. Ku and M. Lee, "TextControlGAN: text-to-image synthesis with controllable generative adversarial networks," *Applied Sciences*, vol. 13, no. 8, Apr. 2023, doi: 10.3390/app13085098.
- [18] Q. Cheng, K. Wen, and X. Gu, "Vision-language matching for text-to-image synthesis via generative adversarial networks," *IEEE Transactions on Multimedia*, vol. 25, pp. 7062–7075, 2022, doi: 10.1109/TMM.2022.3217384.
- [19] M. Chopra, S. K. Singh, A. Sharma, and S. S. Gill, "A comparative study of generative adversarial networks for text-to-image synthesis," *International Journal of Software Science and Computational Intelligence*, vol. 14, no. 1, pp. 1–12, May 2022, doi: 10.4018/ijssci.300364.
- [20] O. Avrahami, D. Lischinski, and O. Fried, "Blended diffusion for text-driven editing of natural images," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 18187–18197, doi: 10.1109/CVPR52688.2022.01767.
- [21] O. Gafni, A. Polyak, O. Ashual, S. Sheynin, D. Parikh, and Y. Taigman, "Make-a-scene: scene-based text-to-image generation with human priors," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13675, Springer Nature Switzerland, 2022, pp. 89–106.
- [22] Y. Zhou *et al.*, "Towards language-free training for text-to-image generation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2022, vol. 2022-June, pp. 17886–17896, doi: 10.1109/CVPR52688.2022.01738.
- [23] A. Abu-Srhan, M. A. M. Abushariah, and O. S. Al-Kadi, "The effect of loss function on conditional generative adversarial networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 6977–6988, Oct. 2022, doi: 10.1016/j.jksuci.2022.02.018.
- [24] X. Huang *et al.*, "Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM," *Energy*, vol. 246, May 2022, doi: 10.1016/j.energy.2022.123403.
- [25] H. Li, W. Qian, R. Nie, J. Cao, and D. Xu, "Siamese conditional generative adversarial network for multi-focus image fusion," *Applied Intelligence*, vol. 53, no. 14, pp. 17492–17507, Jan. 2023, doi: 10.1007/s10489-022-04406-2.






## BIOGRAPHIES OF AUTHORS






**Marzhan Turarova**    in 2008, she graduated from the D. Serikbayev East Kazakhstan State Technical University, specialty 5B070400 - “Computer engineering and software” (bachelor's degree), in 2010, specialty 6M070400 - “Computer engineering and software” (master's degree). In 2023, she defended her dissertation on the specialty 6D070400 - “Computer engineering and software” and received the degree of Doctor of Philosophy (Ph.D.). Currently, she is an acting associate professor of the Department of Computer and Software Engineering at the L. N. Gumilyov Eurasian National University. She is the author of more than 18 scientific papers, including 2 articles in the Scopus database. Research interests - algorithms and software for solving problems of electrical resistivity tomography, geolocation, image processing, big data analysis, computer vision. She can be contacted at email: Marzhan\_08@mail.ru.






**Roza Bekbayeva**    graduated from the Semipalatinsk State University named after Shakarim with a degree in Automation of Technological Processes in 2000. In 2010 she defended her dissertation in the specialty “05.18.12 - Processes and apparatuses of food production” and received a Ph.D. She began her career in 2001 as an assistant at the Department of Automation and Control of SSU named after Shakarim. Currently, she is an Acting Associate Professor at the Department of Automation, Information Technology and Urban Development of Non-profit limited company Semey University named after Shakarim. She is the author of more than 60 scientific papers, including 1 monograph, 3 provisional patents of the Republic of Kazakhstan for an invention, and 2 articles in the Scopus database. Scientific interests - modeling in the field of engineering and technology for the development of products based on renewable systems. She can be contacted at email: rbekbayeva@internet.ru.






**Lazzat Abdykerimova**    from 1990 to 1995 she worked as a teacher of physics and computer science at Almaty State University named after Abay. In 2011-2013 M.Kh. Taraz Regional University named after Dulati, 6M060200 Master of Natural Sciences “Informatics”. She is currently working at Taraz Regional University named after M.H. Dulaty, Kazakhstan. Currently he is a senior lecturer at the Department of Information Systems. She can be contacted at email: Lazzat\_abdykerim@mail.ru.






**Murat Aitimov**    Ph.D in Philosophy of Instrumentation. Dr. Aitimov Murat is an accomplished scholar with a rich academic background. He currently serves as the Director of the Kyzylorda Regional Branch at the Academy of Public Administration under the President of the Republic of Kazakhstan. With an impressive 28 years of experience in both scientific research and pedagogy, Dr. Murat is a recognized authority in his field. His contributions to academia are significant, as evidenced by his substantial body of work. He has authored 55 scientific articles, a testament to his dedication to advancing knowledge. Notably, 7 of his articles have been featured on Scopus, highlighting their impact and relevance in the global research community. In addition to his articles, Dr. Murat has authored three influential books. He also holds an innovation patent in the realm of instrumentation and information systems, showcasing his multidimensional expertise. His expertise and insights continue to drive progress in the field of instrumentation, making him a valuable resource for researchers and academics alike. For those seeking to connect with Dr. Aitimov Murat, he can be reached via email at murat.aytimov@mail.ru.






**Aigulim Bayegizova**    candidate of physical and mathematical, assistant professor. Currently working at the L.N. Gumilyov Eurasian National University at the Department of Radio Engineering, Electronics and Telecommunications. Has more than 40 years of scientific and pedagogical experience, publishing articles in the Scopus database. publication of an article in the Scopus database. She can be contacted at email: baegiz\_a@mail.ru.






**Ulmeken Smailova**    is an experienced senior manager of the “Center of Excellence” of the Autonomous Educational Organization “Nazarbayev Intellectual Schools”, located on Hussein bin Talal Street in Astana, Kazakhstan (postal code: 010000). Candidate of Physical and Mathematical Sciences, 13/05/16 - Application of computer technology, mathematical research and mathematical methods in scientific research. Associate Professor in the specialty 05.13.00 - “Informatics, computer technology and management.” Associate Professor in the specialty 05.13.00 - “Informatics, computer technology and management.” Has more than 70 scientific and methodological articles. She can be contacted at email: samilova\_tarsu@mail.ru.



**Leila Kassenova**    graduated from Al-Farabi Kazakh State University in 1992 with a degree in Physics. Since 2011 - Candidate of Pedagogical Sciences. Since 2021 - Associate Professor in the specialty “Physics”. Currently, she is an Associate Professor of the Department of Information Systems and Technologies at Esil University (Astana). She is the author of more than 60 scientific papers, including 2 textbooks, 1 monograph, 5 articles in the Scopus database (h=3). Research interests – nanostructural analysis, artificial intelligence, machine learning. She can be contacted at email: kassenovalejla@gmail.com.



**Natalya Glazyrina**    graduated from the Institute of Automation, Telecommunications and Information Technology of Omsk State Transport University in 2003. She defended her doctoral thesis on Computer Technology and Software in 2015. From 2016 to the present time, she is an associate professor of the Department of Computer and Software Engineering at L.N. Gumilyov Eurasian National University. She has authored more than 50 papers. Her research interests include mathematical and computer modeling, artificial intelligence, automation of technological processes. She can be contacted at email: glazyrina\_ns\_1@enu.kz.