Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 1, October 2023, pp. 494∼505
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i1.pp494-505

Medical variational autoencoder and generative adversarial network for medical imaging

Zakaria Rguibi[1], Abdelmajid Hajami[1], Dya Zitouni[1], Yassine Maleh[2], Amine Elqaraoui[1]
[1]Research Laboratory Watch Laboratory for Emerging Technologies (LAVETE), Hassan First University of Settat, Settat, Morocco
[2]Ecole Nationale des Sciences Appliquées (ENSA), Sultan Moulay Slimane University, Beni Mellal , Morocco

Article Info

Article history:

Received Apr 19, 2023
Revised Jul 9, 2023
Accepted Jul 16, 2023

Keywords:

Generative adversarial network
Generative models
Medical data
Medical imaging
Variational autoencoder

ABSTRACT

Generative adversarial networks have succeeded promising results in the medical imaging field. One of the most significant challenges in this regard is the lack of or limited data sharing. In our work, an approach for combining generative adversarial network (GAN) and variational autoencoder (VAE) models has been proposed to improve the accuracy and efficiency of medical image analysis tasks. Our approach leverages the capacity of VAEs to acquire condensed feature representations, and the ability of GANs to generate high-quality synthetic images to learn an embedding that keeps high-level abstract visual qualities. Inception score (IS) and Fréchet inception distance (FID) score have been generated in order to demonstrate the high quality of images. Based on the score results, our approach demonstrates the potential of VAE-GAN fusion models and clearly outperforms existing methods on a variety of medical image analysis tasks. The suggested algorithm is explained, as are the results and evaluations.

Corresponding Author:

Zakaria Rguibi
Research Laboratory Watch Laboratory for Emerging Technologies (LAVETE)
Hassan First University of Settat
Settat 21000, Morocco
Email: rguibi.fst@uhp.ac.ma

1. INTRODUCTION

The scarce accessibility of medical data for researchers is a result of legal and regulatory measures that safeguard individual privacy. This occurs because personal medical information is considered extremely sensitive and confidential [1]. The scarcity of data poses a significant challenge for researchers, who rely on extensive datasets to test hypotheses and develop novel treatments and therapies [2]. Without access to real-world data, researchers may struggle to grasp the full extent of a medical condition or identify trends and patterns critical for new treatment development [3]. One potential solution is generating data that mimics real-world data without being based on actual individuals. This can be achieved using simulations or other methods that enable researchers to study complex systems without real-world data [4].

Although not a perfect substitute for real-world data, this approach can offer valuable insights and help researchers better understand various medical conditions' underlying mechanisms [5]. It can also bridge data gaps and provide researchers with larger datasets, even if not based on real individuals [6]. Researchers face significant challenges due to the lack of medical data resulting from laws and regulations restricting the public sharing of private information [7]. By generating data resembling real-world data, researchers can overcome these obstacles and continue making substantial contributions to the medical field [7].

GANs, or generative adversarial networks, are a class of deep learning models that can create new data samples resembling those in a training dataset [8]. They have found applications in various domains, such as image, text, and music generation. In the field of medical imaging, GANs have been employed for several tasks, including generating synthetic medical images, enhancing the quality of low-resolution images, and imputing missing or corrupted image data.

Many papers have been published in recent years on topics including such GAN-based image synthesis, image translation, and image super-resolution in medical imaging, indicating that there is a growing body of literature on the use of GANs in medical imaging. The researchers in [9]-[11] demonstrate the potential of GANs to improve the accuracy and efficiency of medical image analysis tasks, while also highlighting the challenges and limitations of using GANs in this context. Several studies like [11]-[15] illustrate the utilization of GANs in the field of medical imaging, underscoring their potential to enhance both the precision and effectiveness of tasks related to medical image analysis.

Variational autoencoders (VAEs) [16] are a type of generative model that uses a combination of neural networks and statistical methods to learn a compact representation of a dataset, referred to as the latent space, from which novel data samples can be produced. VAEs were first introduced in 2013 by Kingma and Welling [17], and have since been widely used in a variety of applications, including image generation, text generation, and representation learning. VAEs are based on the concept of learning a distribution across data, rather than utilizing a deterministic mapping from input to output. This enables VAEs to capture the diversity of the data and generate novel samples that resemble the training set. VAEs consist of two networks: a decoder network that converts the latent representation back to the original input domain, and an encoder network that associates each input data sample with a latent representation. Generally, the feedforward neural networks composing the encoder and decoder networks are trained to optimize the likelihood of the training data.

The following papers [18]-[20] are examples of research that has explored the use of VAEs in various applications. These papers show some of the various applications of VAEs in a variety of domains, as well as the potential of these models for unsupervised learning and generative modeling. Combining VAE and GAN models might increase the efficiency and adaptability of generative models while also enabling new applications in a range of fields. The remainder of the essay is structured as follows: An overview of relevant studies on the topic is given in section 2. The difficulties of implementing VAE using GAN models are covered in section 3. We provide a summary of the suggested approach and its execution in section 4. The topic and prospective paths for this study are covered in detail in section 5. Finally, in section 6, we wrap up our research and list its main results.

Our unique contribution in this paper is two-fold. Firstly, we propose an architecture that integrates the strengths of both GANs and VAEs, thus effectively mitigating the individual limitations of these two advanced methods. This integrated approach aims to enhance the process of generating synthetic medical data, thereby addressing the scarcity issue that has long hampered progress in the field. Secondly, we provide a comprehensive evaluation of our proposed algorithm using various scores, which stands to substantiate our approach's efficacy and reliability. By successfully addressing the limitations of the existing models, we hope our research will lead to significant advancements in medical research, particularly in areas where the accessibility of real-world data is limited or restricted. This effort not only pushes the boundaries of synthetic data generation but also opens new avenues for researchers to gain deeper insights into various medical conditions and ultimately drive innovation in treatment developments.

2. RELATED WORKS

A form of the generative model called variational autoencoder-generative adversarial networks (VAE-GANs) combines the benefits of VAEs and GANs. A number of papers on VAE-GANs have been published, exploring various aspects of these models, such as their ability to generate high-quality synthetic data, learn robust and interpretable latent representations, and their potential for use in a variety of applications, such as image generation, text generation, and data augmentation. Among the remarkable publications in this field is the original VAE-GAN paper by Larsen *et al.* [21], which introduced the VAE-GAN model and demonstrated its effectiveness on several benchmark datasets. Then, the paper by Zhang *et al.* [22], proposed an improved VAE-GAN model that used a hierarchical latent structure to better capture the structure of the data. Other works have focused on improving the stability and performance of VAE-GANs, such as the use of adversarial training techniques by Meschederet *et al.* [23] or the incorporation of auxiliary tasks by Li *et al.* [24]. VAE-GANs have shown promise as a powerful tool for generative modeling and have been applied to a wide range of tasks and domains. There are numerous original papers in the medical imaging field that demonstrate the application of VAE-GAN [25], which introduces a unique VAE-GAN architecture. Dozens of trials have been carried out to establish its effectiveness. Trial outcomes indicate that this VAE-GAN model outperforms previous state-of-the-art arterial spin labeling (ASL) image synthesis approaches, with an accuracy improvement of up to 42.41% in dementia detection tasks after adding synthesized ASL pictures from the new model. Mostapha *et al.* works [26] is also one of the recent works that use VAE-GAN architecture, in this study, a generative adversarial network using a 3D variational autoencoder is proposed as a semi-supervised out-of-sample detection approach (VAE-GAN).

Cackowski *et al.* [27], suggests a deep-learning method for quick and flexible MR picture harmonization (ImUnity). A VAE-GAN network, along with a confusion module, an optional biological preservation module, picture contrast modifications, and several 2D slices from different anatomical regions in each patient in the training database, is employed for self-supervised training. VAE-GAN algorithms being used in the medical industry [28], [29].

3. CHALLENGES OF DEPLOYING VAE WITH GAN

VAEs outperform other generative models in several ways, including the ability to learn a continuous and structured latent space and handle complex data distributions. They do, however, have a few limitations, including the need for careful design and optimization, as well as the possibility of mode collapse, in which the model only learns to generate a limited number of modes or types of data. Despite these limitations, VAEs continues to be a popular and effective tool for generative modeling, as well as an active field of research in deep learning [30], [31].

VAEs and GANs are both generative models signify their ability to create additional data instances that closely resemble an existing training data set. VAEs are a type of deep learning model that uses a combination of neural networks and statistical methods to learn a compact representation of a dataset, known as the latent space, from which new data samples can be generated. GANs, on the other hand, take a different approach, involving two neural networks that compete to produce the most realistic-looking data samples [32].

Combining VAE and GAN models has the potential to produce more diverse and high-quality generated samples, as the VAE's ability to learn a latent space representation can be combined with the GAN's ability to generate realistic-looking data. However, training and optimizing such a model can be difficult because the two types of models can have competing objectives and require different training methods [33]. Research on the combination of VAE and GAN models is ongoing and has the potential to enhance the quality and variety of produced data samples. As previously stated, combining VAE and GAN models can be challenging since these two types of models have different objectives and require different training methods. The following are some of the specific challenges that may arise when deploying a VAE-GAN fusion model:

- The goals of VAEs and GANs are similar; VAEs are taught to increase the probability of the training data, whereas GANs are trained to reduce the difference between the produced samples and the training data. Aligning these goals can be challenging since increasing the probability of training data can lead to overfitting, while reducing the gap between produced samples and training data can lead to underfitting.
- It can be challenging to train both the VAE and the GAN simultaneously since they are typically trained using different methods and optimization algorithms. For this reason, it may be necessary to employ alternate training methods, such as training the VAE and GAN in tandem or using a shared latent space representation for both models.
- Handling mode collapse: one of the difficulties in training GANs is the possibility of mode collapse, which occurs when a model can only produce a small number of modes or types of data rather than capturing the full diversity of the training data. This is particularly difficult in a VAE-GAN fusion model, as the GAN's tendency to mode collapse can compromise the extent to which the training data's distribution can be fully captured by the VAE.
- Optimizing the hyperparameters: the size and structure of the neural networks, the learning rate and optimization algorithm, and the amount of regularization are all hyperparameters that can affect the performance of VAEs and GANs. Finding the optimal values for these hyperparameters can be difficult, while good performance on a given dataset may necessitate careful tuning and experimentation.

Deploying a VAE-GAN fusion model can be challenging since it requires careful design and optimization to achieve good performance.

4. METHOD

Combining GANs with VAEs models offers the potential to increase the quality and diversity of samples generated while also addressing some of the limitations of each of these models separately [34]. Some of the potential advantages of VAE-GAN fusion models are as follows:

- Improved generation quality: by combining VAEs ability to learn a compact latent space representation with GANs ability to generate high-quality synthetic samples, VAE-GAN fusion models can produce diverse and realistic-looking generated samples. This is especially useful for applications such as image generation, where the quality of the generated samples is crucial.
- Improved feature learning: whereas GANs are taught to reduce the difference between the produced samples and the training data, VAEs are often trained to increase the probability of the training data. Combining these two goals might help VAE-GAN fusion models develop more insightful and understandable feature representations, which can be beneficial for tasks like representation learning and unsupervised learning.
- Better handling of complex data distributions: VAEs can capture the full diversity of the training data and handle complex data distributions that other generative models may find difficult to model. Combining VAEs and GANs can improve GANs' ability to handle complex data distributions and generate more diverse and high-quality samples.
- Mode collapse, which occurs when a GAN model only learns to create a small number of modes or types of data instead of capturing the complete diversity of the training data, is a prevalent problem. By introducing a VAE into the GAN architecture, it may be able to decrease the incidence of mode collapse and enhance the model's capacity to capture all of the variety of training data.

Combining VAE and GAN models could boost the efficacy and adaptability of generative models, allowing for usecases in many fields.

4.1. Variational autoencoder

Mathematically, VAE algorithms are based on the concept of variational inference, which is a method of approximating complex probability distributions with simpler distributions. The goal of VAE analysis is to learn a simpler distribution that closely approximates the true distribution of the data, which is often complex and unknown. VAEs employ a dual-component model consisting of an encoder network and a decoder network to accomplish this. The encoder network transforms an input data sample into a parameter vector representing a distribution in the latent space or a latent representation. Conversely, the decoder network is trained to reconstruct the input data sample from the latent representation, mapping it back to the initial input space [16].

The key mathematical intuition underlying VAEs is that the encoder network's latent representation can be used to approximate the true distribution of the data and that the decoder network can be used to generate new samples that are similar to the training data. The training objective, which typically involves maximizing the likelihood of the training data while also regularizing the latent representation to encourage it to be compact and structured, determines the quality of the approximated distribution and the generated samples. Some of the key equations that define the VAE algorithm are as follows.

Using the (1), the encoder network maps an input data sample x to a latent representation z:

$$z = f(x) \tag{1}$$

where f is the encoder network.
The decoder network restores the original input space from the latent representation z, using the (2):

$$x' = g(z) \tag{2}$$

where g is the decoder network.
The encoder and decoder networks are trained to maximize the likelihood of the training data, using the following objective function as (3):

$$L(x, x') = logp(x|x') \tag{3}$$

where x is the input data sample and x' is the reconstructed data sample.

To regularize the latent representation and encourage it to be compact and structured, VAEs typically use a prior distribution p(z) over the latent space, and minimize the Kullback-Leibler divergence between the approximated distribution q(z—x) and the prior distribution, using the following objective function as (4).

$$D_K L(q(z|x)||p(z)) \tag{4}$$

The objective function for a VAE is typically a combination of the likelihood and regularization objectives [35], and is given by (5).

$$L = E[L(x, x')] - D_K L(q(z|x)||p(z)) \tag{5}$$

These equations define the basic principles and operations of the VAE algorithm and provide a mathematical foundation for understanding how VAEs create fresh samples that are comparable to the training data and how they learn a condensed and understandable representation of the data.

4.2. Generative adversarial networks

To create a generative model for a dataset, generative adversarial networks, a subset of deep learning models, merge neural networks with game theory. GANs were first introduced by Goodfellow *et al.* in 2014 [8] and have since gained popularity for various applications such as image, text, and audio generation. GANs consist of two distinct components, including a generator network that is responsible for generating new data samples and a discriminator network that is trained to differentiate between real and generated samples. These two networks are trained together in a competitive manner, with the generator attempting to fool the discriminator and the discriminator trying to accurately classify real and generated samples. Through this training process, the generator network can ultimately generate high-quality synthetic samples that are similar to the training data.

The key mathematical intuition behind GANs is that the generator network learns a mapping from a latent space to the original data space and that the discriminator network provides feedback on the realism of the generated samples. This allows GANs to learn a generative model of the data, and to produce synthetic samples that are both diverse and realistic-looking. GANs have several advantages over other generative models [36], such as the ability to handle complex data distributions and the ability to generate high-resolution images. However, they also have several limitations, such as the potential for mode collapse and the difficulty of training the generator and discriminator networks simultaneously [37]. Despite these limitations, GANs remain a popular and effective tool for generative modeling and continue to be an active area of research in the field of deep learning. Here are some of the key equations that define the GAN algorithm.

The generator network maps a latent representation z to a data sample x, using the (6):

$$x = G(z) \tag{6}$$

where G is the generator network.
The discriminator network maps a data sample x to a classification score s, using the (7):

$$s = D(x) \tag{7}$$

where D is the discriminator network.

The discriminator and generator networks are trained concurrently, with the generator network trying to maximize the classification score of the generated samples, and the discriminator network trying to minimize the classification score of the generated samples. This is achieved using the following objective functions as (8) and (9).

$$L_G = E_{z\ p(z)}[log(1 - D(G(z)))] \tag{8}$$

$$L_D = E_{x\ p_data(x)}[log(D(x))] + E_{z\ p(z)}[log(1 - D(G(z)))] \tag{9}$$

In the GAN training process, both the generator and discriminator networks are trained simultaneously using their respective objective functions, denoted as L_G and L_D. The prior distribution across the latent area is depicted as p(z), and the genuine data distribution is expressed as p_data(x). Data instances and latent representations are symbolized by x and z, correspondingly.

The objective function for a GAN is typically a combination of the generator and discriminator objective functions, and is given by (10):

$$L = L_G + L_D \tag{10}$$

The (10) define the basic principles and operations of the GAN algorithm and provide a mathematical foundation for understanding how GANs learn a generative model of the data, and how they generate high-quality synthetic samples that are similar to the training data.

4.3. The proposed algorithms

Our model aims to generate synthetic samples that closely resemble the training data using a GAN after a VAE learns a compact and structured representation of medical imaging data. The VAE's encoder network is responsible for transforming input data samples into a latent representation, while the decoder network is tasked with converting the latent representation back into the original input space. The latent representation from the VAE is then utilized to train the GAN to create synthetic samples that resemble the training data. By combining the VAE and GAN, this model can develop a generative model of medical imaging data that can be used to create new samples and perform unsupervised learning tasks. The Algorithm 1 describes the basic operations of the VAE-GAN model:

Algorithm 1 Training loop of VAE/GAN model

$\theta_{Enc}, \theta_{Dec}, \theta_{Dis} \leftarrow$ network parameters Initialization
for i=1, 2, .., number of epochs **do**
  $X \leftarrow$ select mini batch
  $Z \leftarrow Enc(X)$
  $L_{prior} \leftarrow D_K L(q(Z|X)||p(Z))$
  $\tilde{X} \leftarrow Dec(Z)$
  $L_{llike}^{Dis_l} \leftarrow -E_{q(Z|X)}[p(Dis_l(X)|Z)]$
  $Z_p \leftarrow$ sample from N(0, 1)
  $X_p \leftarrow Dec(Z_p)$
  $L_{GAN} \leftarrow log(Dis(X)) + log(1 - Dis(\tilde{X})) + log(1 - Dis(X_p))$
  $\theta_{Enc} \leftarrow \theta_{Enc} - \nabla_{\theta_{Enc}}(L_{prior} + L_{llike}^{Dis_l})$
  $\theta_{Dec} \leftarrow \theta_{Dec} - \nabla_{\theta_{Enc}}(\lambda L_{llike}^{Dis_l} - L_{GAN})$
  $\theta_{Dis} \leftarrow \theta_{Dis} - \nabla_{\theta_{Dis}} L_{llike}^{Dis_l} - L_{GAN}$
end for

The basic training process for the VAE-GAN model is described in this algorithm, which includes steps for initializing the networks, sampling data from the training dataset, passing the data through the encoder and decoder networks, computing the loss functions for the VAE and GAN, and updating the network weights. The algorithm returns the trained VAE-GAN model, which can be used to generate synthetic samples and perform unsupervised learning tasks. The encoder in our model is a convolutional neural network consisting of three convolutional layers and one fully connected layer. The first convolutional layer has 64 filters, each with a 5x5 size and a stride of 2, and employs zero padding to maintain the input's dimensions. The second and third convolutional layers have 128 and 256 filters, respectively, and share the same filter size, stride, and padding as the first layer. The output of the third convolutional layer is flattened and passed through a fully connected layer with 2,048 units.

The fully connected layer's output is then fed into two additional fully connected layers, each containing 128 units. These layers correspond to the mean and log variance of the latent representation. The latent representation is then sampled from a normal distribution using the learned mean and variance. Our model's decoder is a deconvolutional neural network with four deconvolutional layers and one fully connected layer. The decoder takes the latent representation from the encoder and passes it through a fully connected layer with 8x8x256 units, where it is reshaped into a 256x8x8 tensor. Then, four deconvolutional layers with the same number of filters, filter size, stride, and padding as their corresponding convolutional layers in the encoder are applied to this tensor. To ensure pixel values fall within the range [-1, 1], a tanh activation function is applied to the output of the final deconvolutional layer, resulting in a 1x64x64 tensor.

The discriminator component of our model is a convolutional neural network with four convolutional layers and one fully connected layer. It takes an image as input and processes it through four convolutional layers, which have the same filter size, stride, and padding as the corresponding deconvolutional layers in the decoder. The output of the last convolutional layer is flattened and passed through a fully connected layer with a single unit, generating a scalar value indicating the likelihood that the input image is real.

The primary purpose of our model is to train a VAE-GAN model. The VAE encodes input images into a latent space and decodes the latent vectors back into images. The GAN uses a discriminator network to determine whether an image is real or generated. The VAE and GAN in the model are trained together, with the GAN providing additional regularization for the VAE. As a result, the model can generate high-quality images similar to the training data. The model was trained on chest X-ray images and is intended for use in generating medical imaging.

Our VAE-GAN algorithm generates synthetic chest X-ray images. The model comprises an encoder, decoder, and discriminator. The encoder is a convolutional neural network that maps an X-ray image to a latent representation, which is a compact, low-dimensional representation of the image. The decoder, a deconvolutional neural network, takes the latent representation as input and maps it back to an image in the original space. The discriminator is a convolutional neural network that takes an image as input and outputs a scalar value representing the likelihood that the image is real (i.e., not generated by the decoder). While the encoder and decoder aim to generate images similar to the actual ones, the discriminator seeks to differentiate between real and synthetic images.

The objective of the VAE-GAN model is to generate synthetic images that closely resemble real ones by learning a mapping from the original X-ray image space to the latent space. This is achieved by minimizing both the adversarial loss and the reconstruction loss, which represents the discrepancy between the original image and the one produced by the decoder (i.e., the difference between the probability assigned by the discriminator to the real image and the synthetic image). In summary, our algorithm is a VAE-GAN model that is designed to generate synthetic X-ray images of the chest. The encoder, decoder, and discriminator in the model are trained to reduce adversarial loss and reconstruction loss.

4.4. Datasets and hyper parameters

Our dataset consisted of 1,341 medical images sourced from National Institutes Of Health Chest X-Ray Dataset [38] contains over 100,000 chest X-ray images, annotated with various pathologies. The dataset covers a wide range of patient demographics, ensuring the generalizability of our results. The images were preprocessed by resizing them to a standard resolution of 256x256 pixels, followed by normalization to the range [0, 1] without any data augmentation techniques.

In the decoder, our models use backward convolution (also called fractional striding) with a stride of 2 to increase the resolution of the original image. By switching the direction of the convolution, we can perform backward convolution and employ striding for upsampling. We used the RMSProp algorithm for model training, setting the learning rate to 0.0001 and the batch size to 32.

It is worth noting that 1,341 images is a relatively small dataset for training a machine learning algorithm, particularly in the domain of medical imaging, where large amounts of data are frequently required to capture the full range of variability in the data. As a result of the limited size and diversity of the training data, the algorithm may have limited performance or generalizability. This is a common challenge in medical imaging, and it can be addressed using methods like data augmentation, transfer learning, and active learning. These challenges highlight the strength of our approach to generating good, realistic images from small amounts of data.

5. RESULTS AND DISCUSSION

A statistic for assessing the effectiveness of generative models, including VAE-GAN models, is the inception score [39]. It evaluates the produced pictures' quality and variety. More realistic and varied pictures are being produced by the model, as indicated by a higher inception score. The specifics of your model and the data that it was trained on will dictate how the inception score should be interpreted. The inception score formula is as (11).

$$IS = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log p(y_i|x_i)\right) \tag{11}$$

In (11), IS denotes the inception score, N signifies the total number of images assessed, x_i refers to the ith image under evaluation, and y_i is the class label predicted by the inception model for the ith image. The inception model is a pre-trained deep convolutional neural network employed to assess the generated images' quality and diversity. The probability $p(y_i|x_i)$ illustrates the inception model's confidence in predicting the class label y_i for the image x_i. The inception score is computed as the exponential of the average log probabilities of the class labels predicted by the inception model for the generated images. Our VAE-GAN model was trained on a dataset of 1,341 X-ray images. We found the following results Table 1 after evaluating the model using the inception score.

Table 1. Comparison of IS and standard deviations at different training epochs for the VAE-GAN fusion model

Epoch vs score	Mean IS	Standard deviation
Real data	**1.8542**	**0.4173**
Epoch 1	1.1037	0.0459
Epoch 100	1.6008	0.2334
Epoch 350	1.7742	0.4090
Final epoch	**1.8431**	**0.4186**

This indicates that the model was able to generate high-quality and diverse images, as a generated data inception score closer to that of real data corresponds to more realistic and diverse generated images. In conclusion, these findings show the power of our VAE-GAN model. The second aspect of evaluating our model is The Fréchet inception distance (FID) [40], which is a common metric for evaluating the performance of generative models including VAE-GAN models. It is a measure of the difference in distribution between real and generated images in an Inception network's feature space. As indicated by the significant decrease in the FID score from 172,662 in the first epoch to 0.1334 in the final epoch (see Figure 1). This strongly suggests that as training progresses, the model will be able to generate more realistic and diverse images.

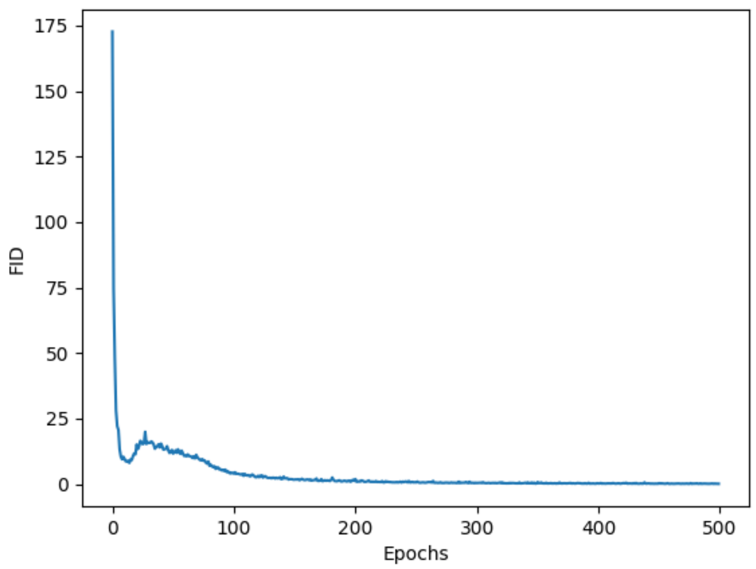

Figure 1. The FID score during training

A lower FID score signifies better model performance. A score of 0 implies that the real and generated images share an identical distribution, while higher scores indicate increasing divergence between the distributions. An inception network is employed to extract features from a collection of generated images and a set of real photographs prior to calculating the FID score. The mean and covariance of both distributions are derived from these features. The FID score is then computed using the squared Euclidean distance between the means of the two distributions and the trace of the product of their covariances.

The most important factor that supports our model is its fast convergence, demonstrating the importance of the VAE block in allowing the GAN model to be faster and more stable during the training process. Furthermore, we used our VAE-GAN model to generate several images as part of our evaluation. Figure 2 show the generated images. These images shows the variety and quality of our model's images. The high level of realism in these images further supports the effectiveness of our model in generating realistic images.

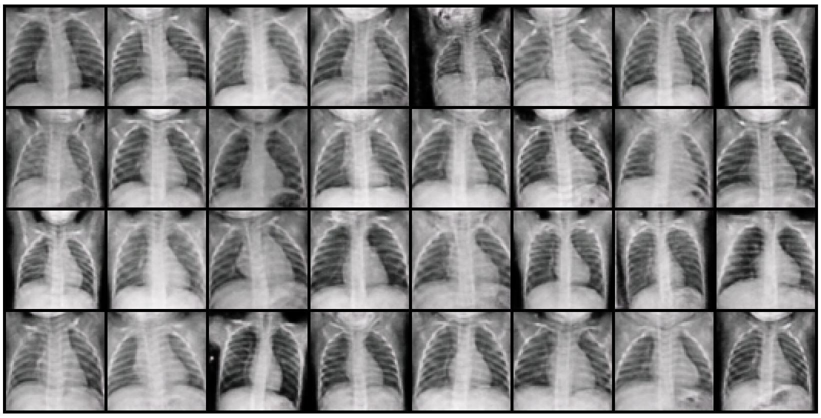

Figure 2. Generated images final

Our research work has been successful in combining VAE and GAN models to produce data that are of great quality and realistic in appearance. This is a promising result, and it suggests that the combination of VAE and GAN models can be a powerful approach for generating synthetic images and for performing unsupervised learning tasks in the domain of medical imaging. Our method effectively produces top-notch artificial images by utilizing a compact dataset of just 1,341 X-ray pictures, which is a remarkable outcome. As previously noted, the restricted quantity and variety of training data can pose challenges for machine learning techniques within the medical imaging field, and attaining excellent performance with smaller datasets can be a tough task.

However, it is important to note that the quality and generalizability of the results obtained with a machine learning algorithm are determined by many factors, including the design and architecture of the algorithm, the quality and diversity of the training data, the effectiveness of the training procedure, and the performance of the algorithm on a held-out test set. Looking to the future, there are many potential directions for further research and development in this area. Some possible directions include: investigating different variations and architectures for the VAE and GAN models, and exploring how these different models affect the quality and diversity of the generated images. Developing new training algorithms and techniques that can improve the stability and convergence of VAE-GAN models, and that can enable the models to handle larger and more complex datasets. Applying VAE-GAN models to different types of medical imaging data, such as X-ray, magnetic resonance imaging (MRI), or ultrasound images, and studying how the models perform on these different data types. Investigating the potential applications of VAE-GAN models in medical imaging, such as in data augmentation, anomaly detection, or image synthesis.

The capacity of VAE-GAN models to produce realistic-looking, high-quality synthetic pictures is one of its main features. By combining the generative modeling capabilities of GANs with the structured and compact latent representations of VAEs, VAE-GAN models are able to capture complex patterns and structures in the data and to generate realistic-looking images that can be used for various applications. Another important aspect of VAE-GAN models is their ability to perform unsupervised learning. By training the VAE and GAN components jointly, VAE-GAN models are able to learn a generative model of the data without the need for labeled examples or other forms of supervision. This makes VAE-GAN models well-suited for tasks such as data augmentation, anomaly detection, and unsupervised representation learning.

In the context of medical imaging, VAE-GAN models have the potential to offer several benefits. For example, VAE-GAN models could be used to augment small or limited datasets with synthetic images, This can enhance the performance of subsequent machine-learning tasks like segmentation or classification. Additionally, VAE-GAN models could be used to identify abnormal or anomalous images, which could be useful for detecting rare or unusual medical conditions. In our work, we are tackling some of the challenges and limitations of VAE-GAN models including the need for large amounts of computational resources (by using just 341 X-ray images), the potential for mode collapse or other training instability issues (as provided by FID figure our architecture is more stable and converge quickly), and the difficulty of interpreting and understanding the learned latent representations (by using VAE to reinforce the GAN model). These challenges are addressed through research and development in areas such as optimization algorithms [41], regularization techniques [42], [43], and latent space analysis [44]. As a conclusion to our discussion, VAE-GAN models are a promising approach for generating synthetic images and for performing unsupervised learning tasks in the domain of medical imaging. By combining the strengths of VAEs and GANs, VAE-GAN models are able to learn complex and structured generative models of medical imaging data, which can be used for a wide range of applications.

6. CONCLUSION

We have successfully merged VAE and GAN models in the realm of medical imaging to create high-quality synthetic pictures that are equivalent to actual ones. Such encouraging results show that combining VAE and GAN models can be a useful method for unsupervised learning tasks in medical imaging, such as synthetic picture synthesis. Future research and development might take many exciting turns. Future research directions for VAE-GAN models in medical imaging could include looking into different model architectures and training algorithms, applying VAE-GAN models to different types of medical imaging data, and looking into the potential applications of VAE-GAN models in medical imaging tasks. There are various intriguing potentials for continued VAE-GAN model research and development in the realm of medical imaging, and it will be interesting to see how these models evolve and improve in the upcoming years.

REFERENCES

[1] M. A. Rothstein, “Is deidentification sufficient to protect health privacy in research?,” *The American Journal of Bioethics*, vol. 10, no. 9, pp. 3–11, Sep. 2010, doi: 10.1080/15265161.2010.494215.

[2] T. B. Murdoch and A. S. Detsky, “The inevitable application of big data to health care,” *JAMA*, vol. 309, no. 13, p. 1351, Apr. 2013, doi: 10.1001/jama.2013.393.

[3] E. H. Shortliffe and J. J. Cimino, Eds., *Biomedical informatics*. London: Springer London, 2014. doi: 10.1007/978-1-4471-4474-8.

[4] T. Ching *et al*., “Opportunities and obstacles for deep learning in biology and medicine,” *Journal of The Royal Society Interface*, vol. 15, no. 141, p. 20170387, Apr. 2018, doi: 10.1098/rsif.2017.0387.

[5] A. Esteva *et al*., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[6] M. Ghassemi, T. Naumann, P. Schulam, A. L. Beam, I. Y. Chen, and R. Ranganath, “A review of challenges and opportunities in machine learning for health,” *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, vol. 2020, pp. 191–200, 2020.

[7] A. L. McGuire and R. A. Gibbs, “No longer de-identified,” Science, vol. 312, no. 5772, pp. 370–371, Apr. 2006, doi: 10.1126/science.1125339.

[8] I. Goodfellow *et al*., “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[9] D. Nie *et al*., “Medical image synthesis with context-aware generative adversarial networks,” in *Medical Image Computing and Computer Assisted Intervention MICCAI 2017. MICCAI 2017*, Cham: Springer, 2017, pp. 417–425. doi: 10.1007/978-3-319-66179-7_48.

[10] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification,” *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.

[11] D. Mahapatra, B. Bozorgtabar, S. Hewavitharanage, and R. Garnavi, “Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis,” in *Medical Image Computing and Computer Assisted Intervention MICCAI 2017. MICCAI 2017*, Cham: Springer, 2017, pp. 382–390. doi: 10.1007/978-3-319-66179-7_44.

[12] A. Royer *et al*., “XGAN: unsupervised image-to-image translation for many-to-many mappings,” in *Domain Adaptation for Visual Understanding*, Cham: Springer International Publishing, 2020, pp. 33–49. doi: 10.1007/978-3-030-30671-7_3.

[13] H. Li *et al*., “DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019*, Cham: Springer, 2019, pp. 795–803. doi: 10.1007/978-3-030-32251-9_87.

[14] M. Mardani *et al*., "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 167–179, Jan. 2019, doi: 10.1109/TMI.2018.2858752.

[15] A. Lahiri, V. Jain, A. Mondal, and P. K. Biswas, "Retinal vessel segmentation under extreme low annotation: a gan based semi-supervised approach," in *2020 IEEE International Conference on Image Processing (ICIP),* Oct. 2020, pp. 418–422. doi: 10.1109/ICIP40778.2020.9190882.

[16] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019, doi: 10.1561/2200000056.

[17] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," Dec. 2013, [Online]. Available: http://arxiv.org/abs/1312.6114.

[18] M. Mirza and S. Osindero, "Conditional generative adversarial nets," Nov. 2014, [Online]. Available: http://arxiv.org/abs/1411.1784.

[19] G. Kwon, C. Han, and D. Kim, "Generation of 3D brain MRI using auto-encoding generative adversarial networks," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019*, Springer, 2019, pp. 118–126. doi: 10.1007/978-3-030-32248-9_14.

[20] H. Yang *et al*., "Synthesizing multi-contrast MR images via novel 3D conditional variational auto-encoding GAN," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 415–424, Feb. 2021, doi: 10.1007/s11036-020-01678-1.

[21] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.09300.

[22] T. Zhang, Z. Yang, and D. Li, "Stochastic simulation of deltas based on a concurrent multi-stage VAE-GAN model," *Journal of Hydrology*, vol. 607, p. 127493, Apr. 2022, doi: 10.1016/j.jhydrol.2022.127493.

[23] L. Mescheder, S. Nowozin, and A. Geiger, "Adversarial variational Bayes: unifying variational autoencoders and generative adversarial networks," Jan. 2017, [Online]. Available: http://arxiv.org/abs/1701.04722.

[24] J. Li, C. Niu, and K. Xu, "Learning part generation and assembly for structure-aware shape synthesis," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11362–11369, Apr. 2020, doi: 10.1609/aaai.v34i07.6798.

[25] F. Li, W. Huang, M. Luo, P. Zhang, and Y. Zha, "A new VAE-GAN model to synthesize arterial spin labeling images from structural MRI," *Displays*, vol. 70, p. 102079, Dec. 2021, doi: 10.1016/j.displa.2021.102079.

[26] M. Mostapha *et al*., "Semi-supervised VAE-GAN for out-of-sample detection applied to MRI quality control," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019*, Cham: Springer, 2019, pp. 127–136. doi: 10.1007/978-3-030-32248-9_15.

[27] S. Cackowski, E. L. Barbier, M. Dojat, and T. Christen, "ImUnity: a generalizable VAE-GAN solution for multicenter MR image harmonization," *Medical Image Analysis*, vol. 88, p. 102799, Aug. 2023, doi: 10.1016/j.media.2023.102799.

[28] X. Liu *et al*., "Dual-cycle constrained bijective vae-gan for tagged-to-cine magnetic resonance image synthesis," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2021, pp. 1448–1452. doi: 10.1109/ISBI48211.2021.9433852.

[29] A. Volokitin *et al*., "Modelling the distribution of 3D brain MRI using a 2D slice VAE," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020,* Cham: Springer, 2020, pp. 657–666. doi: 10.1007/978-3-030-59728-3_64.

[30] X. Chen, N. Pawlowski, M. Rajchl, B. Glocker, and E. Konukoglu, "Deep generative models in the real-world: an open challenge from medical imaging," Jun. 2018, [Online]. Available: http://arxiv.org/abs/1806.05452

[31] D. Xue *et al*., "Advances and challenges in deep generative models for de novo molecule generation," *WIREs Computational Molecular Scienc*e, vol. 9, no. 3, May 2019, doi: 10.1002/wcms.1395.

[32] R. Gieselmann and F. T. Pokorny, "Standard deep generative models for density estimation in configuration spaces: a study of benefits, limits and challenges," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2020, pp. 5238–5245. doi: 10.1109/IROS45743.2020.9340994.

[33] L. Mi, M. Shen, and J. Zhang, "A probe towards understanding GAN and VAE models," Dec. 2018, [Online]. Available: http://arxiv.org/abs/1812.05676

[34] S. W. A. Canchumuni, J. D. B. Castro, J. Potratz, A. A. Emerick, and M. A. C. Pacheco, "Recent developments combining ensemble smoother and deep generative networks for facies history matching," *Computational Geosciences*, vol. 25, no. 1, pp. 433–466, Feb. 2021, doi: 10.1007/s10596-020-10015-0.

[35] D. Shen, C. Qin, C. Wang, H. Zhu, E. Chen, and H. Xiong, "Regularizing variational autoencoder with diversity and uncertainty awareness," Oct. 2021, doi: 10.24963/ijcai.2021/408.

[36] H. GM, M. K. Gourisaria, M. Pandey, and S. S. Rautaray, "A comprehensive survey and analysis of generative models in machine learning," *Computer Science Review*, vol. 38, p. 100285, Nov. 2020, doi: 10.1016/j.cosrev.2020.100285.

[37] A. Srivastava, L. Valkov, C. Russell, M. U. Gutmann, and C. Sutton, "VEEGAN: reducing mode collapse in GANs using implicit variational learning," May 2017, [Online]. Available: http://arxiv.org/abs/1705.07761

[38] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 3462–3471. doi: 10.1109/CVPR.2017.369.

[39] S. Barratt and R. Sharma, "A note on the inception score," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.01973

[40] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.08500

[41] V. Leyton-Ortega, A. Perdomo-Ortiz, and O. Perdomo, "Robust implementation of generative modeling with parametrized quantum circuits," *Quantum Machine Intelligence,* vol. 3, no. 1, p. 17, Jun. 2021, doi: 10.1007/s42484-021-00040-2.

[42] S. Sinha and A. B. Dieng, "Consistency regularization for variational auto-encoders," May 2021, [Online]. Available: http://arxiv.org/abs/2105.14859

[43] Z. Xiao *et al*., "Improving transferability of adversarial patches on face recognition with generative models," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 11840–11849. doi: 10.1109/CVPR46437.2021.01167.

[44] A. Vahdat, K. Kreis, and J. Kautz, "Score-based generative modeling in latent space," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.05931

BIOGRAPHIES OF AUTHORS

Zakaria Rguibi a leading artificial intelligence researcher from Morocco, serves as the Head of the Artificial Intelligence Department at PROCHECK SA. He is also pursuing a Ph.D. in computer science at Université Hassan 1er, focusing on machine learning, deep learning, and medical imaging. His work spans developing machine learning models for fraud detection, designing natural language processing systems, and applying AI algorithms to optimize manufacturing processes. He is a member of the International Advisory Committee for AI for global goals, a United Nations initiative leveraging AI to achieve sustainable development goals. He can be contacted at email: rguibi.fst@uhp.ac.ma.

Professor Abdelmajid Hajami a distinguished scholar in informatics and telecommunications, was born in Morocco in 1975. He earned his Ph.D. from Mohamed V Souissi University and currently serves as a Professor at the Faculty of Science and Technology in Settat, Morocco. Additionally, he is the Director of the LAVETE Laboratory, where his research focuses on telecommunications and artificial intelligence. His other research interests include wireless networks, elearning, and biotechnology communications. He can be contacted at email: abdelmajid.hajami@uhp.ac.ma.

Dya Zitouni holds a Ph.D. degree and is currently a distinguished Professor at the Faculty of Sciences and Technologies, University Hassan First, Settat, Morocco. As a respected member of the LA-VETE Laboratory, Zitouni's research is primarily focused on advancing knowledge in the realm of artificial intelligence, with particular emphasis on machine learning and deep learning applications. This pursuit has led Zitouni to contribute significantly to the growing field of AI, fostering innovation and inspiring the next generation of AI researchers and practitioners. He can be contacted at email: zitouni.dya@uhp.ac.ma.

Dr. Yassine Maleh is an associate professor of cybersecurity and IT governance at Sultan Moulay Slimane University in Morocco. He is the founder of the IEEE Consultant Network Morocco and the African Research Center of Information Technology & Cybersecurity. A senior member of IEEE and other professional organizations, He has published extensively in information security, IoT security, and IT governance. He serves in editorial roles for several journals and on committees for numerous international conferences. He has also been recognized as a Publons Top 1% reviewer in 2018 and 2019. He can be contacted at email: assine.maleh@ieee.org.

Amine Elqaraoui is a dedicated Ph.D. student at the LAVETE Laboratory, Hassan 1st University in Settat, Morocco. His research focus is on reinforcement learning within the field of artificial intelligence. He is also a Member of the Artificial Intelligence Department at PROCHECK SA working on many projects. He can be contacted at email: a.elqaraoui@uhp.ac.ma.