Journal of Basrah Researches (Sciences) **50**(2), 20 (2024)
DOI: https://doi.org/10.56714/bjrs.50.2.3

High-Resolution Guided Image Synthesis Using Latent Diffusion Model Technology

Sura Ahmed Abd, Khawla Hussein Ali*, Zaid Ameen Abduljabbar

Department of Computer Science, College of Education for Pure Sciences, University of Basrah, Iraq

A R T I C L E I N F O

Received 14 April 2024
Revised 15 August 2024
Accepted 29 August 2024
Published 31 December 2024

K e y w o r d s :

Diffusion Models, Image Synthesis, Reverse Diffusion, Forward Process. Ultrasound Images.

Citation: S. A. Abd et al., J. Basrah Res. (Sci.) **50**(2), 20 (2024).
DOI:https://doi.org/10.56714/bjrs. 50.2.3

A B S T R A C T

We introduce the latent diffusion model used in medical ultrasound image synthesis. We point out that precision is the issue, and installing ultrasound images was completed with an accuracy of 97.47% since ultrasound demands greater accuracy. It has some particular disadvantages because it operates in real-time and requires operator settings. Considering these challenges, our model has a lot of promise to provide accurate and lifelike ultrasound images. Even though it is hard to calculate the precise answer for this optimization, applying the backpropagation method merely once can produce an approximation. In order to train a diffusion model with the value and outcomes (FID: 2.870, CLIP: 0.209, SSIM: 0.9923, and LPIPS: 0.92) that we promised, we generated synthetic images of roughly 300 ultrasound images. were acquired. Expanding the use of artificial intelligence in medical imaging is the aim of this endeavor. Since this is a novel problem, the study will serve as a foundation and source of inspiration for researchers looking into possible applications of diffusion models in medical image production. The URL https://www.kaggle.com/datasets/suraahmed56/computer-vision-medical-images provides access to synthetic images.

1. Introduction

Diffusion Models with latent variables motivated by ideas in nonequilibrium thermodynamics are probabilistic models [1],[2],[3],[4]. Diffusion Models provide high-quality image synthesis findings. Diffusion Models most outstanding performance was achieved using a weighted variational bound during training, which was made possible by a new relationship between denoising score matching and diffusion probabilistic models

*Corresponding author email : Khawla.ali@uobasrah.edu.iq

©2022 College of Education for Pure Science, University of Basrah. This is an Open Access Article Under the CC by License the CC BY 4.0 license.

N: 1817-2695 (Print); 2411-524X (Online)
ine at: https://jou.jobrs.edu.iq

.

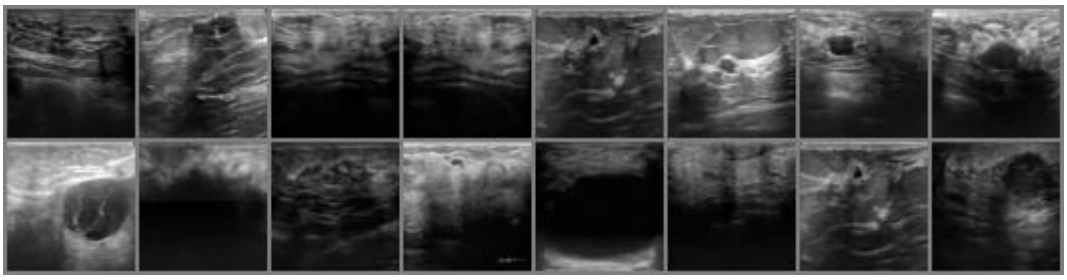

Fig. 1. Selected samples from our best Image model.

The public's attention has been drawn to guided image synthesis with user notes due to large-scale models of language-image (LLI), which have recently been developed [5],[6],[7]. By merging unsupervised learning with text-based training learning from a reference image, usually a painting with thick outlines.

New users have control over the contents of the final image [8]. The user can choose the final scene's coloring combination with the attached rough painting outlines, while the text timely ascertains the general semantics of the image[9]. Examine the information challenge in a novel and captivating way [10]. Training these models aims to remove noise from noisy inputs [11]. When an input is fully noisy, and no information from the input is retained, a DM creates a new sample[12]. However, when a DM is applied to a partially noisy input, some of the noisy input's information is preserved, and the denoising process is conditioned on it. The model attempts to reproduce the input[13]. The amount of noise could be viewed as a variable information congestion[14],[15]. Unlike a property of the trained model itself, the issue during model inference is something we can alter externally, such as the size of the latent space in an autoencoder [16]. This work investigates several diffusion model topologies and how well they produce artificial images that closely mimic accurate data.

High-resolution, photorealistic image production is still a significant challenge in image synthesis and computational creativity. Generative models (LDMs) and conventional convolutional neural networks (CNNs) have shown tremendous progress in image-generating challenges. However, when scaling up to high-resolution outputs, these methods frequently run into problems, including artifacts that take away from the photorealism of the synthesized images, loss of detail, and unnatural textures. Furthermore, these models may offer little control over the production process, making it challenging to direct the synthesis process per particular user requirements or to alter specific elements of the created images without degrading the overall coherence and quality. Latent Diffusion Models (LDMs) have become a viable solution to these problems by fusing the effectiveness of latent space representations with the advantages of diffusion models. Nevertheless, the use of LDMs in guided, high-resolution image synthesis presents a unique set of issues, such as the requirement for significant computational resources, the difficulty of preserving high fidelity to the guiding input in the latent space, and the challenge of making sure that the generated images are realistic and diverse while abiding by predetermined guidelines or constraints.

This research aims to tackle these problems by developing advanced techniques and methodologies for high-resolution guided image synthesis using latent diffusion model technology. (Section 3.2) specifically, seeks to enhance the capability of LDMs to generate photorealistic images at high resolutions, improve the efficiency and scalability of the synthesis process, and refine the ability to control and guide the image generation process. The goal is to enable the production of high-quality, high-resolution images that meet specific user-defined criteria, opening up new possibilities for applications in areas such as digital art, entertainment, and virtual reality.

We also discuss potential applications of these techniques to typical application issues such as data scarcity and motion artifacts. We also investigate the possible uses of synthetic images in teaching and research. The investigation's findings provide insight into the bright future of model diffusion in image synthesis. With an emphasis on applying diffusion models, this study offers multiple essential contributions to image synthesis. The following are the research's contributions:

- Diffusion models are frequently employed in the field of technology adoption research to examine how new technologies are embraced by people, groups, or society. For companies, legislators, and designers of technology, this information is essential.
- Policy Development: Diffusion models are used to help create policies that aim to encourage and make it easier for positive innovations to be adopted in areas like public health, education, and technology. In situations where prompt adoption is essential to the well-being of society, this is especially pertinent.
- Progress in Image Synthesis: By presenting and investigating the use of cutting-edge diffusion models, this work makes a substantial contribution to the field of image synthesis. This gives rise to a novel solution to the age-old problem of safely and effectively obtaining excellent-quality photographs.
- Common Challenges Risk mitigation: The research explores how diffusion models might help reduce common problems such as data shortages and motion artifact impact. In situations where patient mobility or insufficient data are usually issues, it provides ways to enhance image quality.
- Improved Image Fidelity: Diffusion models have an amazing capacity to produce images with extraordinary fidelity, according to the analysis. This helps provide synthetic ultrasound images that are highly similar to real-world pictures, which is advantageous for research and clinical applications.
- Forecast Capabilities: Diffusion models can be used to forecast the future adoption of new inventions by analyzing historical adoption trends and comprehending the relevant elements. To properly plan and strategize, this is useful for corporations, legislators, and consumers.

The paper is organized as follows: Section 2 explains the literature review. Section 3 shows the methodology of the Latent Diffusion Model technique in four sections (3.1. Forward process, 3.2. Additional input z for conditioning, 3.3. Reverse process, and 3.4modified from the trained sampling diffusion model). Section 4 states: Experiments with the DMs. Section 5 shows the discussion and conclusion.

2. Related work

Numerous research studies in this area synthesize and play a vital part in this endeavor by making it possible to create synthetic images equivalent to actual ones. A group of deep generative models called diffusion models has shown great promise and excels in image synthesis[17]. They have advanced computer vision research in many fields, such as enhancing images.

In a paper Evangelos Ntavelis et al. [18] A novel method for producing detailed and static 3D objects is put forth by the authors. A 3D auto-decoder structure forms the basis of their approach. This method also makes use of a 3D auto-decoder that learns embeddings in the latent space. After decoding these embeddings, a volumetric representation can be created, enabling geometric display and a consistent look. The latent space contains embedded features from the target dataset. This makes decoding into 3D representations efficient. Because this approach can handle both solid and articulated objects, it also provides versatility. It can handle a wide variety of datasets because it is sufficiently broad. It's interesting that this strategy can function both with and without camera

information in any case. During training, it effectively picks up camera information as needed. The evaluation findings demonstrate that, on various datasets and benchmarks, the created 3D assets beat state-of-the-art alternatives.

Ling Yang et al. [19] introduce CONPREDIFF, an original method to enhance diffusion-based image synthesis by incorporating context prediction. Following the diffusion denoising blocks of the training stage, we explicitly reinforce each point to use a context decoder to forecast Next, separate the decoder for inference from its neighborhood context (e.g., multi-hop features, tokens, or pixels). Moreover, for this reason, each point may repeat itself more precisely because it maintains its semantic connections to the surrounding context.

Yanyu Li et al.[20] Our generic solution tackles these problems. It is the first of its type, unlocking Diffusion models from textual to image on portable gadgets in under two seconds. We achieve this by refining step distillation and creating a thriving network architecture. In particular, we identify the redundant parts of the original model and use data distillation to reduce the computation of the image decoder to propose an efficient UNet. In addition, we investigate several training approaches and incorporate regularization using classifier-free guidance to improve the step distillation further. Our extensive MS-COCO experiments demonstrate that our model has eight denoising steps.

Weijia Wu et al.[21] Using text-guided image synthesis, this approach expands on the pre-trained diffusion model to generate perceptual data. We demonstrate how a decoder module can efficiently decode the diffusion model's extensive latent code into precise perception annotations. By training the decoder with less than 1% (about 100 images) of manually labeled photos, an infinitely large annotated dataset can be produced. Afterward, these artificial data sets could be used to train various perception models for further tasks. Provide rich and dense pixel-wise label datasets for various downstream tasks, including instance, semantic, and depth estimation to show that the recommended approach is effective. It accomplishes two key goals: (1) it produces cutting-edge outcomes on instance and semantic segmentation, and (2) it is notably more resilient.

An effective alternative that offers a solid framework for the synthesis of high-quality images is the diffusion model. The paper "Cascaded Diffusion Models for High Fidelity Image Generation" examines this innovative technique" paper Jonathan Ho et al.[22] which demonstrates how effectively class-conditional images on the ImageNet dataset may be generated. The paper solves frequent problems such compounding errors throughout the sampling process by providing conditioning augmentation, a novel technique for data augmentation. Using a multi-step super-resolution technique, exhibits greater performance in high-fidelity picture production with improved FID scores and classification accuracy. The outcomes demonstrate notable enhancements in image quality, surpassing cutting-edge models like as BigGAN-deep and VQ-VAE-.

Promising answers are provided by recent developments in generative deep learning, namely in the area of diffusion models. In order to effectively address The paper by Alexander Shmakov et al.[23] A unique method called end-to-end latent variable diffusion models (LVDDMs) is presented, which combines the advantages of diffusion processes with the inverse problems of high-energy physics. latent variable models. provide a new unified architecture called latent variational diffusion models (LVDDMs) that integrates A thorough variable framework with a latent learning about state-of-the-art generative art techniques. We show that this method works effectively for both guaranteeing that the learnt posterior distributions follow established physics constraints and reconstructing global distributions of theoretical kinematic quantities. It is noteworthy that our combined approach offers a distance to the truth, distribution-free, that is roughly twenty times shorter than a third the usual latent diffusion models and less than the most

recent non-latent baseline. The fundamental limitation of the LVDDM technology is the significant computational complexity of the diffusion process. Moreover, it is challenging to ensure that the generated data precisely captures the richness and diversity of the actual environment, which may necessitate further optimization and refinement.

In the paper Shaoyan Pan et al.[24] the authors suggest a novel way to make radiotherapy planning easier: they take routine magnetic resonance imaging (MRI) and use it to generate synthetic computed tomography. In order to avoid the requirement for separate cross-sectional simulation and image registration, the objective is to produce high-quality cross-sectional images using MRI data. During treatment planning, this method lowers the patient's radiation exposure and eliminates ambiguity in the environment. The suggested approach is known as the noise diffusion probabilistic model (MC-DDPM) based on the MRI-to-CT converter. In addition, it improves the diffusion process by converting MRI to high-quality sCT utilizing a moving window transducer network. Two procedures make up this approach. Gaussian noise is introduced in the first step, known as the forward process noisy CT images when compared to actual CT scans.The second is the opposite procedure: noise-free CT scans are produced by isolating the CT noise conditioned by the matching MRI using a variable window V-net (Swin-Vnet). Furthermore, the brain dataset Brain sCTs generated by MC-DDPM yielded state-of-the-art quantitative results, such as a normalized cross-correlation (NCC) of 0.983, a multi-scale structural similarity index (MS-SSIM) of 0.965, and a mean absolute error (MAE) of 43,317 Hounsfield units (HU). We conclude that MC-DDPM enables the development of reliable, high-quality sCT images in a matter of minutes by efficiently capturing the intricate interaction between CT and MRI images. The planning of radiation will be significantly impacted by this revolutionary method.

The authors of the paper Tiange Xiang et al.[25] provide an independent denoising technique specially developed for diffusive denoising generative models in MRI denoising. Where the dispute is raised One crucial medical imaging technology is a form of magnetic resonance imaging (MRI); nevertheless, in order to get a good signal-to-noise ratio (SNR) for MRI, lengthy scan periods are frequently necessary, leading to higher expenses and discomfort for the patient. In order to overcome this difficulty, the authors concentrate on MRI scan noise reduction, which is especially constrained by SNR. The suggested technique combines diffusion models with statistics-based denoising theory, dubbed Diffusive Denoising Models for MRI Denoising (DDM^2). It runs in three stages, using modal generation to cut down on noise. Inference-related input noise measurements are shown as samples from the diffusion Markov chain mean posterior distribution. DDM^2 is supervised Subjective, in contrast to the majority of earlier MRI denoising techniques that depend on supervised training datasets. This is crucial since it is not feasible to gather supervised data for various anatomy, MRI scanners, and scanning parameters. The authors were able to conduct tests on in vivo MRI data sets in order to assess the performance. Superior denoising performance was proven by DDM^2, as assessed by clinically important optical qualitative and quantitative criteria. To sum up, DDM^2 efficiently lowers noise in diffusion MRI scans, improving picture quality and maybe improving patient care in the course of treatment planning.

Authors of the paper Jiahang Cao et al.[26]suggest a novel class of diffusion models based on SNNs. For generative tasks, specifically noise reduction using diffusion MRI images, the aim is to use artificial neural networks' energy efficiency and biological acceptability. Furthermore, the suggested approach, known as Denoising Propagation Probabilistic Models (SDDPM), features an architecture that is solely Spiking U-Net. Similarly, with only 4 time steps, this architecture delivers a large energy reduction and performs similarly to its counterpart in an artificial neural network (ANN). In terms of terms that are generated: Although earlier studies have mostly concentrated on improving SNNs for classification tasks, SDDPM investigates the generative

capabilities of SNNs. It surpasses previous SNN-based methods and produces state-of-the-art outcomes in generative challenges models that generate. The authors' threshold-oriented strategy, which even without training, boosts performance by 2.69%, is a nice idea. In the end, SDDPM opens up new research opportunities and constitutes a significant advancement in the field of SNN-based image production.

These connected articles shed light on the application of deep learning techniques and image synthesis to ultrasonic imaging for medical applications. With an emphasis on employing these models to produce synthetic images, researchers can use these studies to push the frontiers of diffusion model construction and progress. In addition to talking about some of the challenges facing the sector in comparison to. While traditional methods like (GANs) and (VAEs) have considerable potential, they frequently fail to generate high-quality output and capture the intricate geometric elements needed for realistic 3D models.

3. Methodology

The goal of the proposed method is to synthetic medical modalities (Ultrasound images) by using a diffusion model as shown Figure 2.

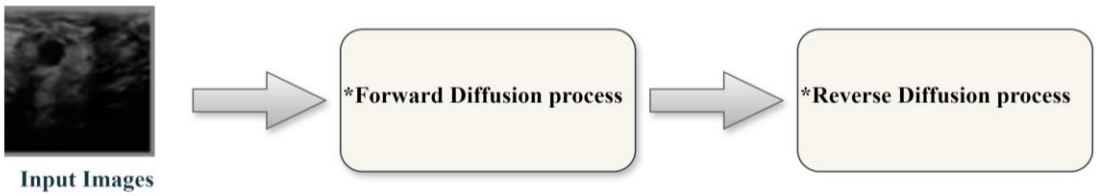

Fig. 2. The general strategy

The suggested approach comprises two processes: a forward process that represents adding noise using a Gaussian noise and a reverse process that represents noise removal as depicted in Figure 3.

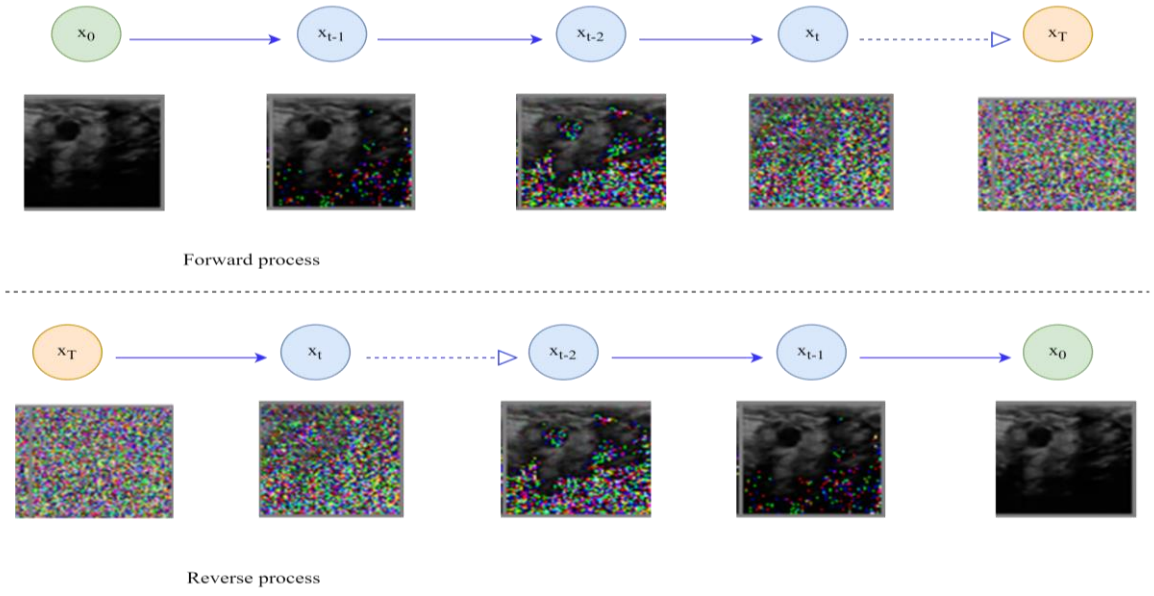

Fig. 3. Schematic of the Proposed Method

While diffusion models are only a few latent variable models, they have many implementation possibilities. The model architecture, the parameterization of the reverse process using the Gaussian distribution, and the variances β_{t-1} of the forward process need to be chosen in (Section 3.2). We present a new explicit relationship (that guides our decision-making) between diffusion models and blurring score matching, leading to a simplified weighted variational bound objective (In Section 3.3) for diffusion models. Finally, objective evidence and simplicity are discussed in Section 4 as reasons for our model design. Equation (4)'s components categorize our conversation:

3.1. Forward process

For our purposes, let's say that an input image, x0 ~ q(x|z), that has been sampled from a data distribution of real samples, q(x), that needs to be simulated, gradually gains noise thanks to a forward process. In this case, the data depends on a variable called z. The extra Gaussian noise at each time-step t, where t ∈ {1, T}, is a Markov chain with T steps and variance. Its sole dependencies are on the training variable and the sample from the previous phase. As a result, the distribution can be expressed as a latent variable. The diffusion forward operation consequently has the following expression:

$$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t \mid \mathbf{x}_{t-1}, z) = \prod_{t=1}^{T} \mathcal{N}\left(\mathbf{x}_t; \mu t = \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \sigma t = \beta_t \mathbf{I}\right) \quad (1)$$

By defining, it is feasible to compute x_t without computing every sample at earlier stages. This allows a sample x_t to be sampled as follows:

$$\mathbf{x}_t \sim q(\mathbf{x}_t \mid \mathbf{x}_{t-1}, z) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha})\mathbf{I}\right) \quad (2)$$

Variance β may be fixed in this work, β improves over T steps from 10−4 to 0.02.

3.2. Additional input z for conditioning

To condition the synthesis y, an encoder τ_γ with weights γ is employed. The U-Net's intermediary layers are mapped to the middleman depiction $\tau_\gamma(z) \in R^{M \times d}$, which is the outcome. of mapping this encoder. Like R, the domain-specific encoder for language prompts may employ a transformer technique. Using a latent representation from a trained auto-encoder, we generate the encoding for higher-dimensional conditioning data, such as a picture. The auto-encoder is implemented in this work as a convolutional neural network with three down- and three up-sampling layers. There are 128 dimensions in the latent representation, and it is mapped to the intermediate layers the U-Net. This provides us with the image encoding and text encoding for sample $\mathbf{x}_t$, conditioned on z.

3.3. Reverse process

The distribution gets closer to an isotropic Gaussian with a big enough T. Therefore, through the reversal of the unit Gaussian distribution noise addition process N (0, I) samples, the data distribution q(x) may be simulated. But in reality, neither the reverse $q(x_{t-1}|x_t, z)$ nor its statistical estimation is known because obtaining any statistical estimate would require knowledge of the data distribution. However, we can use a parameterized function $\epsilon\theta^{(x_t, t, z)}$ to learn how to approximate $q(x_{t-1}|x_t, z)$. This function can be thought of as a sequence of extra-conditioned de-

noising auto-encoders on the time-step t. Parameterizing a Gaussian and then manually removing the predicted Gaussian noise is easier. For sample x_{t-1}, we therefore have:

$$P_\theta(\mathrm{x}_t(\mathrm{x}_t \mid \mathrm{x}_{t-1}, z) = \mathcal{N}\left(\mathrm{x}_{t-1}; \mu_\theta(\mathrm{x}_t, t), \sigma_\theta(\mathrm{x}_t, t)\right) \quad (3)$$

Next, implementing the reverse procedure for every time-step:

$$P_\theta(\mathrm{x}_t(\mathbf{x}_{1:T} \mid \mathbf{x}_0) = P_\theta(\mathrm{x}_T, T, z) \prod_{t-1}^{T} P_\theta(\mathrm{x}_t \mid \mathrm{x}_{t-1}, z) \quad (4)$$

In this instance, the complete procedure is packaged in a de-noising function $\epsilon_\theta\,(x_t, t, z)$, It has been trained to expect an input de-noised representation, in particular x_0, from x_t. We can rewrite this as $\epsilon_\theta\,(x_t, t, \tau\gamma(z))$ to include the encoder that was used to pre-process the conditioning variable. The goal for this may then be reduced to:

$$L_{=} E_{x,\epsilon\sim N(0,1),t,z} \left[\|\epsilon - \epsilon_\theta(x_t, t, z)\|_1^1 \right] \quad (5)$$

We utilize a U-Net to maximize computing efficiency as our denoising procedure in this study and input latent images into the function for training. In diffusion models, the UNet design comprises an up-sampling path for precise localization and a down-sampling path for context capture. This structure makes the model's ability to process data at various resolutions possible, and it is essential for producing detailed results. A distinguishing feature of the UNet architecture is skip connections. They establish connections between the layers in the up-sampling path and those in the lower. Because the UNet is conditioned on noise levels, it can learn a reverse diffusion process, which is crucial for the quality of generated samples in diffusion models. This design aids in maintaining high-resolution details throughout the network. The model starts with a

simple random noise distribution and eventually learns to recover the data distribution via denoising inputs. The training can be done in several ways, including changing the input or using UNet's adaptive normalization layers. The UNet architecture's adaptability allows it to process multimodal inputs efficiently. This feature is used in diffusion models to condition the generation process on extra data, including class labels, text descriptions, or even other images, improving the model's capacity to generate focused and coherent outputs. as shown in the figure:

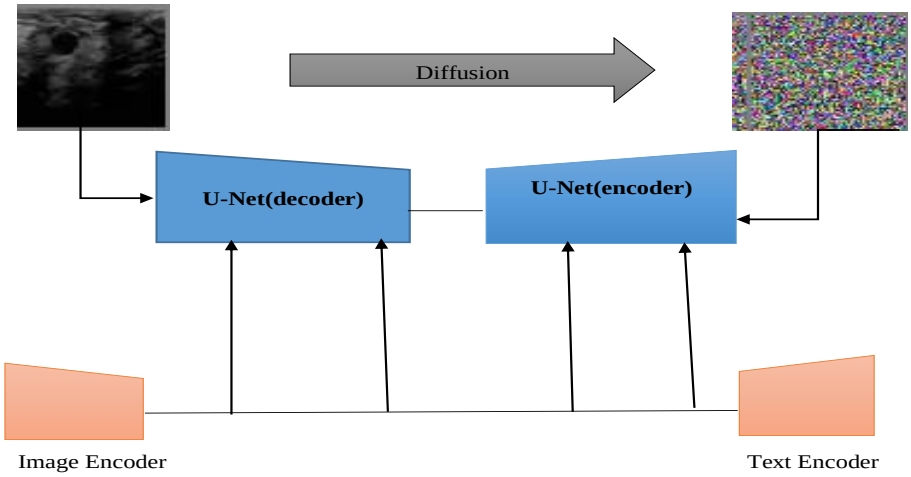

Fig. 4. A summary of the diffusion procedure

3.4 Modified from the trained sampling diffusion model

For taking an example of a image taken from the distribution of learned data and utilizing the supplied conditioning parameters. We use our reverse de-noising function to compute the sample., $x_0 = \epsilon_\theta$ (x_t, t, τγ(z)), and sample $x_T \sim N(0, I)$. Better results are observed in practice, nonetheless, unless noise is reintroduced using the noise schedule step t 1, or using Eq.1. The de-noising function is then used once more to create a latent sample. x_0, to which the schedule is used to add noise until t -2, and so on, until t- t.

4. Experiments

4.1. Dataset

The data set consists of a set of data ultrasound images.
https://www.kaggle.com/datasets/sabahesaraki/breast-ultrasound-images-dataset
split into a train set (1262images) and a test set (317images) (total of 1578 images. Through the following link located in Kaggle. Kaggle features excellent-quality images[27]. This dataset was split into subgroups for testing and training., adhering to the identical data arrangement as in. An open pose is used to extract skeletons. The images differ in terms of background, lighting, and points of view. The test set identities do not overlap concerning the dataset.

4.2. Evaluation Metrics

Four separate Criteria for evaluation are applied to the model. In addition, the Two methods for measuring accuracy are the Structure Similarity Index (the SSIM score)[28] and Trained Perceptual Image Patch Along Clarity (LPIPS)[29]. Comparably, LPIPS computes the perceived difference between the created and reference images, whereas SSIM computes the image at the pixel level. The degree of realism in the produced images can be evaluated using the fractional inception distance (FID) [30]. A comparison is made between the generated image and the ground-truth image distributions, and a Wasserstein-2 distance is calculated. Particulars of Implementation: We used T = 1000 noising steps to train our model[31].

We examine the potential of the suggested scaling factor to consider diffusion models when creating images of varying sizes. Specifically, we substitute the self-attention layer's tuning factor in diffusion models and evaluate their performance on a subset of the dataset [32],[33],[34], which includes, respectively, over 300 image pairs filtered by CLIP, without any training. We randomly choose text-image pairings from each dataset and produce images corresponding to texts using different evaluation methods. We consider trained Latent Diffusion and Stable Diffusion as diffusion models to evaluate the performances[35],[36],[37]. The former, set up to synthesize images with a resolution of 512 × 512, keeps four pixels more than the latter, which has a default resolution of 256 × 256.

Table 1. Latent Diffusion in different resolution settings.

Dataset	FID	CLIP	SSIM	LPIPS	Various evaluation
Ankan et al[38]	6.3671	---	0.7312	0.1678	
Jonathan Ho et al[39]	3.17				IS(9.46±0.11)
Weilun Wang et al[40]	18.8	---	0.422		mIoU(39.2)
Proposed method	2.870	0.209	0.4728	0.92	

Table 1. Table displays the results together using the efficacy at the trained models.

The proposed model is compared quantitatively to various cutting-edge models using Fritz Inception Distance (FID), Visual image patch similarity that is taught (LPIPS) and its Measure of the Structure Similarity Index (SSIM).

It is important here to point out that the synthetic images generated by DMs are in different epochs: the first synthetic image was produced in only 315 epochs, the second in 348 epochs, the third in 392 epochs, and so on, until the final image in Figure 5, which was produced in only 1000. squeeze out's

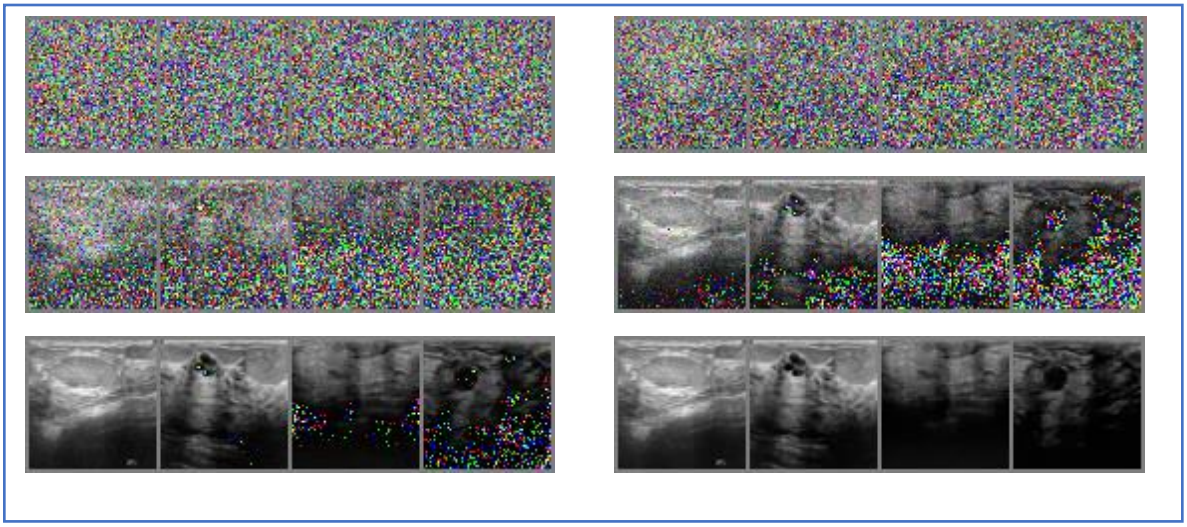

Fig. 5 DMs receiving training to produce artificial CT images across several epochs

The generative path in our suggested denoising diffusion model is dependent on the pose and style. determines the likely transfer trajectories by dividing the issue into a series of diffusion steps that go forward and backward

Table 2. The Ultrasound dataset's count of images and accuracy

Train Loss	Test Loss	Train Accuracy	Test Accuracy
0.0346	0.0868	97.47%,	97.00%

Figure .6 shows the training and test loss curve and training and Test accuracy of the model.

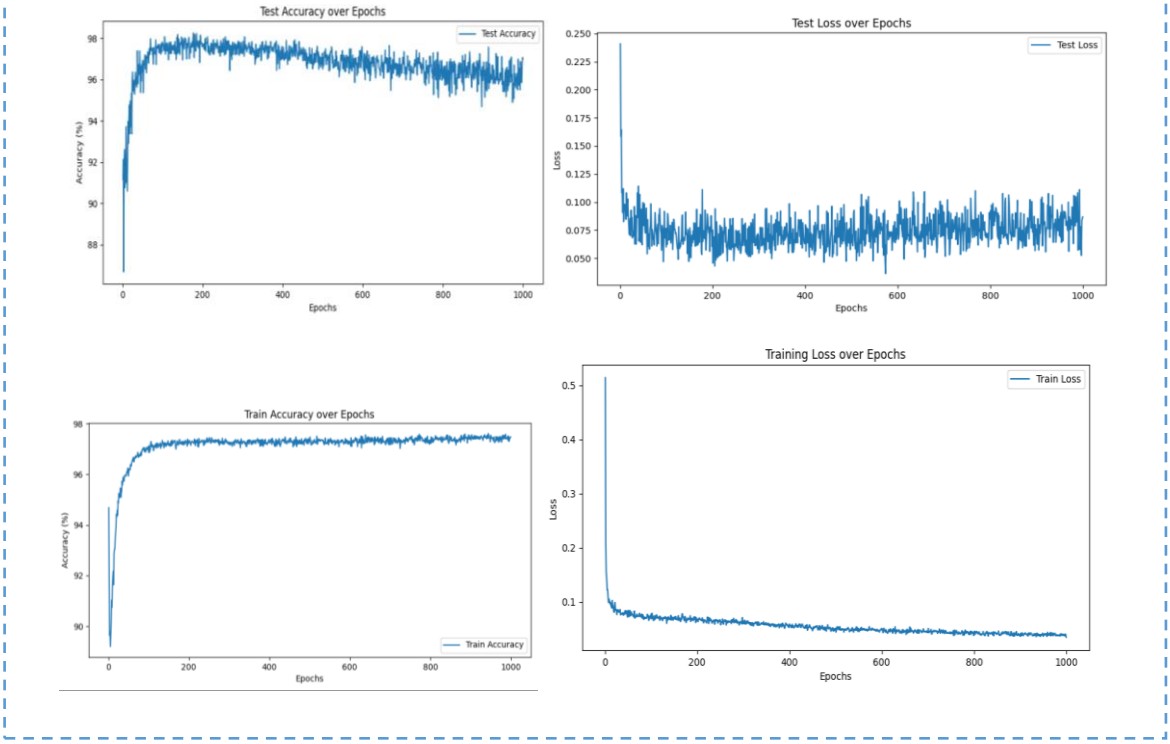

Fig. 6. The accuracy and loss of the ultrasound images

Here, it observe that the test accuracy is 97.00%, indicating that for around 97% of the test samples, the model accurately predicts the category labels. The precision of the training is also very crucial. It gauges how well the training set of data matches your model. With a training accuracy of 97.47%, the model is said to function well with the training set of data. Additionally, the test loss also known as the validation loss determines the degree to which the model's predictions agree with the ground truth labels found in the data. Better performance is indicated by a lower test loss. The test loss has a value of 0.0868. Lastly, loss of training: A model's fit to the training set of data is measured by train loss, also known as training loss. Low train loss is preferred, just like test loss.

In conclusion, it's critical to remember the following points:

- To start with, your model performs well on training and testing sets of data.
- Second, the model performs well as evidenced by the comparatively low loss values.

5. Discussion

The synthesis of ultrasonic images has been carried out with a 97.47% accuracy rate using Latent diffusion models. Ultrasound imaging poses particular unique challenges since it requires operator settings and functions in real-time. Despite these challenges, LDMs show a great deal of promise for obtaining a sufficient degree of precision and generating realistic ultrasound images. Depending on how well they work, LDMs may be useful in creating synthetic ultrasound images for use in training by medical professionals and AI models. Its lower accuracy when compared to other techniques suggests that more improvement and optimization are needed to guarantee that LDMs for ultrasonic imaging can catch the tiny oscillations observed in clinical settings. Because they can produce unseen images, LDMs are therefore preferable over traditional approaches. As a result, our model has performed better than conventional techniques like GANs and VAEs. GAN

models are well known for their potentially unstable training and decreased generational variability since they employ adversarial training. is reliant on a proxy-based loss.

modalities like MRI and CT for a number of reasons related to the constraints and nature of ultrasound imaging.

1. Image diversity and quality: Ultrasound images might not meet the same
2. Tissue properties and comparison: Less contrast: Compared to MRI and CT scans, ultrasound images might occasionally display less color, which makes it more challenging to identify abnormalities and distinguish between various tissue types.
3. Anatomical structures' complexity: One typical application of ultrasound technology is accurate imaging of dynamic and complicated anatomical systems, such as the heart and moving organs.
4. Other disorders: Ensuring that the model appropriately gathers and categorizes all possible abnormalities since a variety of medical diseases can be detected with ultrasound imaging.
5. Model optimization: More intricate model designs and approaches will be required to handle the unique problems that ultrasonic imaging presents. Perhaps we shouldn't use this anymore
6. The general low resolution of ultrasonic image synthesis is responsible for several ultrasound imaging challenges, including poor contrast and picture quality, operator reliance, and complexity of the body structures being viewed. Improved model optimization, data collecting, and contemporary training approaches should be used to overcome these issues.

6. Conclusion

This paper focuses on generating medical images, as it is important to synthesize these images using diffusion models that match reality. This paper specifically examines the synthesis of medical images, and ultrasound imaging, as they appear with the value and results (FID: 2.870, CLIP: 0.209, SSIM: 0.9923, and LPIPS: 0.92). The accuracy was: 97.47%, which is considered excellent, and the extent of the devices' ability LDMs to enhance imaging medical, which contributes to improving medical imaging. The results of this paper confirm that obstetric diffusion models continue to advance, provide solutions to challenges, and open new horizons for developing their ability to adapt and apply them in several therapeutic contexts, such as medical learning and personal support. Diffusion models remain in advanced centers where there is high accuracy and ideal quality in Generating high-resolution indicative image.

References

[1] Z. Jin, X. Shen, B. Li, and X. Xue, "Training-free Diffusion Model Adaptation for Variable-Sized Text-to-Image Synthesis," no. NeurIPS, pp. 1–14, 2023, [Online]. Available: http://arxiv.org/abs/2306.08645

[2] Z. Chen et al., "Hierarchical Integration Diffusion Model for Realistic Image Deblurring,"no. NeurIPS, pp. 1–12, 2023, DOI: http://arxiv.org/abs/2305.12966

[3] J. Singh, S. Gould, and L. Zheng, "High-Fidelity Guided Image Synthesis with Latent Diffusion Models," no. Lli, pp. 5997–6006, 2023, doi: 10.1109/cvpr52729.2023.00581.

[4] L. Sun, J. Chen, Y. Xu, M. Gong, K. Yu, and K. Batmanghelich, "Hierarchical Amortized GAN for 3D High Resolution Medical Image Synthesis," IEEE J. Biomed. Heal. Informatics, vol. 26, no. 8, pp. 3966–3975, 2022, doi: 10.1109/JBHI.2022.3172976.

[5] S. U. Saeed et al., "Bi-parametric prostate MR image synthesis using pathology and sequence-conditioned stable diffusion," pp. 814–828, 2023, DOI: http://arxiv.org/abs/2303.02094

[6] M. J. J. Ghrabat, G. Ma, I. Y. Maolood, S. S. Alresheedi, and Z. A. Abduljabbar, "An effective image retrieval based on optimized genetic algorithm utilized a novel SVM-based

convolutional neural network classifier," Human-centric Comput. Inf. Sci., vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0191-8.

[7] E. T. Khalid, M. Salah Khalefa, W. Yassen, and A. Adil Yassin, "Omicron virus emotions understanding system based on deep learning architecture," J. Ambient Intell. Humaniz. Comput., vol. 14, no. 7, pp. 9497–9507, 2023, doi: 10.1007/s12652-023-04615-8.

[8] X. Xing, C. Wang, H. Zhou, J. Zhang, Q. Yu, and D. Xu, "DiffSketcher: Text Guided Vector Sketch Synthesis through Latent Diffusion Models," no. NeurIPS, 2023, [Online]. Available: http://arxiv.org/abs/2306.14685

[9] Z. A. Abduljabbar et al., "Provably Secure and Fast Color Image Encryption Algorithm Based on S-Boxes and Hyperchaotic Map," IEEE Access, vol. 10, no. February, pp. 26257–26270, 2022, doi: 10.1109/ACCESS.2022.3151174.

[10] D. Podell et al., "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," 2023, [Online]. Available: http://arxiv.org/abs/2307.01952

[11] K. J. Joseph et al., "Iterative Multi-granular Image Editing using Diffusion Models," pp. 8107–8116, 2023, [Online]. Available: http://arxiv.org/abs/2309.00613

[12] A. Vuong et al., "Language-driven Scene Synthesis using Multi-conditional Diffusion Model," no. NeurIPS, 2023, [Online]. Available: http://arxiv.org/abs/2310.15948

[13] J. Kim and H. Park, "Adaptive Latent Diffusion Model for 3D Medical Image to Image Translation: Multi-modal Magnetic Resonance Imaging Study," 2023, [Online]. Available: http://arxiv.org/abs/2311.00265

[14] M. A. Mohammed, M. A. Hussain, Z. A. Oraibi, Z. A. Abduljabbar, and V. O. Nyangaresi, "Secure Content Based Image Retrieval System Using Deep Learning," Basrah Res. Sci., vol. 49.2, no. 2, pp. 94–111, 2023, doi: 10.56714/bjrs.49.2.9.

[15] P. Friedrich, J. Wolleb, F. Bieder, A. Durrer, and P. C. Cattin, "WDM: 3D Wavelet Diffusion Models for High-Resolution Medical Image Synthesis," 2024, [Online]. Available: http://arxiv.org/abs/2402.19043

[16] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," Z. Med. Phys., vol. 29, no. 2, pp. 102–127, 2019, doi: 10.1016/j.zemedi.2018.11.002.

[17] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the Design Space of Diffusion-Based Generative Models," Adv. Neural Inf. Process. Syst., vol. 35, no. NeurIPS, 2022.

[18] E. Ntavelis, A. Siarohin, K. Olszewski, C. Wang, L. Van Gool, and S. Tulyakov, "AutoDecoding Latent 3D Diffusion Models," no. NeurIPS, 2023, [Online]. Available: http://arxiv.org/abs/2307.05445

[19] L. Yang et al., "Improving Diffusion-Based Image Synthesis with Context Prediction," no. NeurIPS, pp. 1–21, 2024, [Online]. Available: http://arxiv.org/abs/2401.02015

[20] Y. Li et al., "SnapFusion: Text-to-Image Diffusion Model on Mobile Devices within Two Seconds," no. NeurIPS 2023, pp. 1–17, 2023, [Online]. Available: http://arxiv.org/abs/2306.00980

[21] W. Wu et al., "DatasetDM: Synthesizing Data with Perception Annotations Using Diffusion Models," no. NeurIPS, pp. 1–13, 2023, [Online]. Available: http://arxiv.org/abs/2308.06160

[22] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded Diffusion Models for High Fidelity Image Generation," Journal of Machine Learning Research, vol. 23, pp. 1–33, 2022.

[23] A. Shmakov, K. Greif, M. Fenton, A. Ghosh, P. Baldi, and D. Whiteson, "End-To-End Latent Variational Diffusion Models for Inverse Problems in High Energy Physics," no. NeurIPS, pp. 1–26, 2023, [Online]. Available: http://arxiv.org/abs/2305.10399

[24] S. Pan et al., "Synthetic CT Generation from MRI using 3D Transformer-based Denoising Diffusion Model.

[26] J. Cao, Z. Wang, H. Guo, H. Cheng, Q. Zhang, and R. Xu, “Spiking Denoising Diffusio Probabilistic Models.” [Online]. Available: https://github.com/AndyCao1125/SDDPM

[27] Z. Lv et al., “Learning Semantic Person Image Generation by Region-Adaptive Normalization,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 3, pp. 10801–10810, 2021, doi: 10.1109/CVPR46437.2021.01066.

[28] K. N. S. Behara, A. Bhaskar, and E. Chung, “Insights into geographical window based SSIM for comparison of OD matrices,” ATRF 2017 - Australas. Transp. Res. Forum 2017, Proc., no. November, pp. 1–6, 2017.

[29] J. Ko, I. Kong, D. Park, and H. J. Kim, “Stochastic Conditional Diffusion Models for Robust Semantic Image Synthesis”.

[30] T. M. R. I. Synthesis, Q. H. Nguyen, and T. Le, “Class Label Conditioning Diffusion Model for Robust Brain Class Label Conditioning Diffusion Model for Robust Brain Tumor MRI Synthesis,” pp. 0–11, 2023, doi: 10.36227/techrxiv.24243829.v1.

[31] M. Antonelli et al., “The Medical Segmentation Decathlon,” Nat. Commun., vol. 13, no. 1, pp. 1–13, 2022, doi: 10.1038/s41467-022-30695-9.

[32] M. J. J. Ghrabat, G. Ma, Z. A. Abduljabbar, M. A. Al Sibahee, and S. J. Jassim, “Greedy Learning of Deep Boltzmann Machine (GDBM)’s Variance and Search Algorithm for Efficient Image Retrieval,” IEEE Access, vol. 7, pp. 169142–169159, 2019, doi: 10.1109/ACCESS.2019.2948266.

[33] Z. A. Abduljabbar et al., “SEPIM: Secure and efficient private image matching,” Appl. Sci., vol. 6, no. 8, 2016, doi: 10.3390/app6080213.

[34] X. Dai et al., “Multimodal MRI synthesis using unified generative adversarial networks,” Med. Phys., vol. 47, no. 12, pp. 6343–6354, 2020, doi: 10.1002/mp.14539.

[35] J. Wolleb, F. Bieder, R. Sandkühler, and P. C. Cattin, “Diffusion Models for Medical Anomaly Detection,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 13438 LNCS, pp. 35–45, 2022, doi: 10.1007/978-3-031-16452-1_4.

[36] P. Dhariwal and A. Nichol, “Diffusion Models Beat GANs on Image Synthesis,” Adv. Neural Inf. Process. Syst., vol. 11, no. NeurIPS 2021, pp. 8780–8794, 2021.

[37] Z. Dorjsembe, H.-K. Pao, S. Odonchimed, and F. Xiao, “Conditional Diffusion Models for Semantic 3D Medical Image Synthesis,” 2023, doi: 10.21227/3ej9-e459.

[38] A. Kumar Bhunia et al., “Person Image Synthesis via Denoising Diffusion Model,” pp. 5968–5976, 2023, doi: 10.1109/cvpr52729.2023.00578.

[39] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” Adv. Neural Inf. Process. Syst., vol. 2020-Decem, no. NeurIPS 2020, pp. 1–25, 2020.

[40] W. Wang et al., “Semantic Image Synthesis via Diffusion Models,” 2022, [Online]. Available: http://arxiv.org/abs/2207.00050

Journal of Basrah Researches (Sciences) **50**(2), 20 (2024)
DOI: https://doi.org/10.56714/bjrs.50.2.3

تركيب صور موجه عالي الدقة باستخدام تقنية نموذج الانتشار الكامن

سرى احمد عبد، خوله حسين علي*، زيد امين عبد الجبار

قسم علوم الحاسوب، كلية التربية للعلوم الصرفة، جامعة البصرة، العراق،

معلومات البحث

الاستلام 14 نيسان 2024
المراجعة 15 أب 2024
القبول 29 أب 2024
النشر 31 كانون الاول 2024

الكلمات المفتاحية

نماذج الانتشار، تركيب الصورة، الانتشار العكسي، العملية الأمامية. صور الموجات فوق الصوتية.

الملخص

نقدم نموذج الانتشار الكامن المستخدم في تركيب صور الموجات فوق الصوتية الطبية. نشير إلى أن الدقة هي المشكلة، وتم الانتهاء من تركيب صور الموجات فوق الصوتية بدقة 97.47% لأن الموجات فوق الصوتية تتطلب دقة أكبر. وله بعض العيوب الخاصة لأنه يعمل في الوقت الفعلي ويتطلب إعدادات المشغل. وبالنظر إلى هذه التحديات، فإن نموذجنا لديه الكثير من الأمل في تقديم صور بالموجات فوق الصوتية دقيقة ونابضة بالحياة. على الرغم من أنه من الصعب حساب الإجابة الدقيقة لهذا التحسين، فإن تطبيق طريقة الانتشار العكسي مرة واحدة فقط يمكن أن يؤدي إلى نتيجة تقريبية. من أجل تدريب نموذج الانتشار بالقيمة والنتائج (FID: 2.870، CLIP: 0.209، SSIM: 0.9923، وLPIPS: 0.92) التي وعدنا بها، قمنا بإنشاء صور اصطناعية لما يقرب من 300 صورة بالموجات فوق الصوتية. تم الحصول عليها. إن توسيع استخدام الذكاء الاصطناعي في التصوير الطبي هو الهدف من هذا المسعى. وبما أن هذه مشكلة جديدة، ستكون الدراسة بمثابة أساس ومصدر إلهام للباحثين الذين يبحثون في التطبيقات المحتملة لنماذج الانتشار في إنتاج الصور الطبية. يوفر عنوان URLhttps://www.kaggle.com/datasets/suraahmed56/computer-vision-medical-images إمكانية الوصول إلى الصور الاصطناعية.

Citation: S. A. Abd et al., J. Basrah Res. (Sci.) **50**(2), 20 (2024). DOI:https://doi.org/10.56714/bjrs. 50.2.3

***Corresponding author email :** Khawla.ali@uobasrah.edu.iq

©2022 College of Education for Pure Science, University of Basrah. This is an Open Access Article Under the CC by License the CC BY 4.0 license.

N: 1817-2695 (Print); 2411-524X (Online)
ine at: https://jou.jobrs.edu.iq