2D reconstruction of ultrasound lung images via diffusion

Ismail Abouamal
Caltech
abouamal@caltech.edu

Shi-Ning Sun
Caltech
ssun3@caltech.edu

Abstract

We present an approach to reconstruct 2D images of lung organs from ultrasound scans using a diffusion model. As a first step, we start by training an unconditional model to generate 2D lung images from complete noise. Subsequently, we extend this methodology to develop a conditional model, utilizing ultrasound scans as measurements to guide the denoising process that generates 2D lung images.

1 Introduction

Background

Ultrasound imaging, also known as sonography, is an important technique used in diverse medical fields. It operates by emitting high-frequency sound waves into the body or surrounding environment, capturing their echoes as they bounce off obstacles and internal structures. These echoes are then processed by sensors to construct detailed images of the surrounding anatomy, including organs, blood vessels, and developing fetuses. In obstetrics, for example, ultrasound serves as a pivotal tool for assessing fetal health, monitoring development, and detecting potential abnormalities. Similarly, in cardiology, it offers a non-invasive means to evaluate the heart's structure and function with remarkable precision, sparing patients the invasiveness of surgical procedures.

Among its many applications, ultrasound finds a prominent role in abdominal imaging. By employing sonography to examine organs like the liver, clinicians can diagnose an array of conditions such as cysts and tumors. Notably, ultrasound presents several advantages over alternative imaging modalities such as X-rays or MRI. It offers a cost-effective alternative to MRI scans while eliminating the need for radiation exposure, making it a safer option for patients. Furthermore, it elevates the necessity of injecting contrast agents, ensuring a comfortable imaging experience for the patient.

Despite its versatility and efficacy, ultrasound encounters challenges in imaging certain organs, particularly the lungs. The technique's effectiveness is notably hindered by its limited ability to penetrate air-filled spaces. In lung imaging, this limitation becomes pronounced as ultrasound waves encounter difficulty traversing through the air-filled lung tissue essential for respiration. Consequently, visualizing deep structures beyond the lungs presents a challenge. The air within the lungs acts as a formidable barrier, impeding the ultrasound waves from reaching and visualizing deeper structures within the chest cavity.

Despite the challenges posed by the air-filled nature of the lungs, ultrasound continues to hold promise as a valuable diagnostic tool, driving ongoing research and advancements aimed at overcoming its limitations and expanding its utility in pulmonary care. The recent progress in machine learning techniques to solve inverse problems suggests a new avenue to overcome the challenges posed by imaging lung tissue with ultrasound. In this report, we discuss one possible approach that looks promising: using a diffusion model to obtain more accurate 2D lung images from ultrasound data.

CS165 winter, 2024.

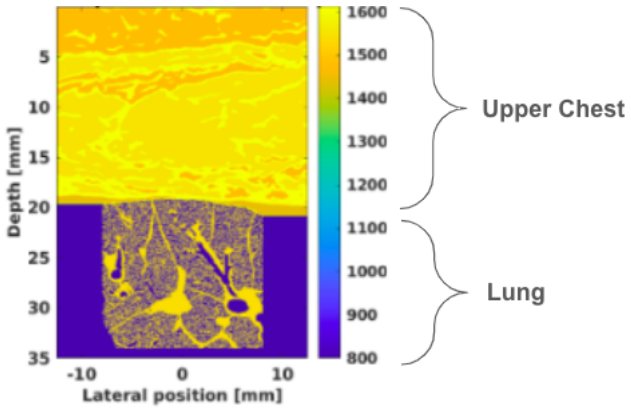

Figure 1: Speed of sound vs depth. The waves travel slower and dissipate at higher rate as they penetrate deeper into the chest.

Research problem

The importance of our research lies in the fact that accurately discerning the shape and structure of internal organs from ultrasound scans is crucial for healthcare practitioners. This capability makes it possible to conduct thorough visual inspections and detect early anomalies such as tumors, enabling timely intervention and treatment.

As discussed in the introduction, the task of accurately delineating the shape of the lung organ using ultrasounds poses a significant challenge. Our research proposes a potential solution to this problem.

We aim to develop and implement the technique of diffusion and guided diffusion to augment the capabilities of ultrasound imaging, thereby enhancing its diagnostic utility. By leveraging this new technology, we seek to overcome the inherent limitations of traditional ultrasound methods. More precisely, our aim is to employ the recent progress in diffusion models to reconstruct the two-dimensional image of a lung organ from the ultrasound scans.

Research Goal

Our project will consist of two steps. We will first train an unconditional diffusion model using our 2D lung images as the training data. Once we obtain this model and are able to generate new 2D lung images with this model, we will train a conditional model using the ultrasound measurements as a guidance of the reverse diffusion process.

2 Background

Diffusion models are a family of generative AI models inspired by non-equilibrium thermodynamics. Roughly speaking, the underlying structure consists of a Markov chain of diffusion steps where each step adds random noise to the data. In what follows, we review the general theory behind these models. For the detail, we refer the reader to [1].

Consider a data point $\mathbf{x_0}$ sampled from an unknown distribution $q(\mathbf{x})$ and define the forward diffusion process as follows. Let $T \in \mathbb{N}$ and $\beta_1, \ldots, \beta_T$ such that $\beta_t \in (0,1)$ for $0 < t \leq N$ and the consider forward process defined by:

$$\mathbf{x}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

That is,

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t).$$

We can check that

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, \tag{1}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t$.

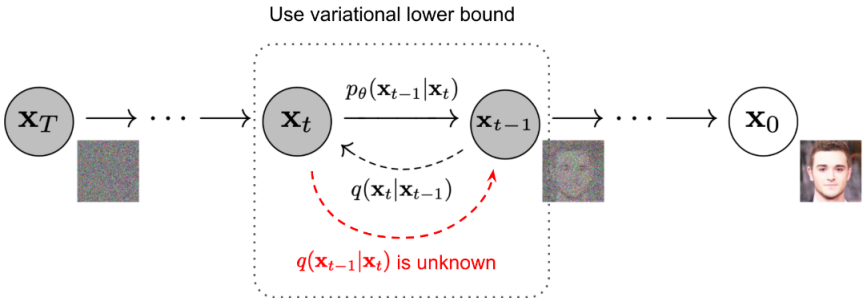

Figure 2: The Markov chain of forward/ reverse diffusion process of generating a sample by slowly adding/removing noise. (Image source: Lilian Weng Blog)

If we know the the conditional distributions $q(\mathbf{x}_{t-1}|\mathbf{x_t})$, we should be able to sample from the distribution from a given noisy sample $\mathbf{x_T} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Unfortunately, this will require the knowledge of the true distribution of $q(\mathbf{x})$. On the other hand, we will now show that the reverse conditional probability distribution is tractable if we condition on $\mathbf{x}_0$. In fact we have

$$
\begin{aligned}
q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) &= q(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_0) \frac{q(\mathbf{x}_{t-1} \mid \mathbf{x}_0)}{q(\mathbf{x}_t \mid \mathbf{x}_0)} \\
&= \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\overline{\alpha}_{t-1}}\right) x_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t} x_t + 2\frac{\sqrt{\overline{\alpha}_{t-1}}}{1-\overline{\alpha}_{t-1}} x_0\right) x_{t-1} + C(\mathbf{x}_t, \mathbf{x}_0)\right)\right)
\end{aligned}
$$

After some algebra, we find that

$$
q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) \sim \mathcal{N}(\tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{1})
$$

with

$$
\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_\mathbf{t}
$$

and

$$
\tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1 - \overline{\alpha}_{t-1})}{1 - \overline{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}} \beta_t}{1 - \overline{\alpha}_t} \mathbf{x}_0.
$$

Using equation equation 1, we find that

$$
\tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \overline{\alpha}_t}} \epsilon_t)
$$

Our goal is to start with a Gaussian noise $\mathbf{x}_T$ and sample from $q(\mathbf{x}_{t-1}|x_t)$. The goal of this model is learn a model p_θ to approximate thesse conditional distributions.

We will assume that reverse process is also normally distributed, that is

$$
p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}(\mu_\theta(\mathbf{x}_t), \Sigma_\theta(\mathbf{x}_t)).
$$

Note that this assumption has deep theoretical foundations from the theory of stochastic differential equations. We then have

$$
p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t).
$$

The learning objective of this model is to minimize the cross-entropy between the underlying distribution $q(\cdot)$ and the learned distribution $p_\theta(\cdot)$. Hence

$$
\begin{aligned}
-\mathbb{E}_{q(\mathbf{x}_0)} \log p_\theta(\mathbf{x}_0) &= -\mathbb{E}_{q(\mathbf{x}_0)} \log \left(\int p_\theta(\mathbf{x}_{0:T}) d\mathbf{x}_{1:T} \right) \\
&\leq -\mathbb{E}_{q(\mathbf{x}_{0:T})} \log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \quad \text{(Jensen's inequality)} \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T})} \left[D_{KL}(q(\mathbf{x}_T|\mathbf{x}_0) \,||\, p_\theta(\mathbf{x}_T)) + \sum_{t=2}^{T} D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \,||\, p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)) - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1) \right]
\end{aligned}
$$

where D_{KL} denotes the KL divergence (also known as conditional entropy). The above model is trained by optimizing the right hand side (also known as the variational bound on the negative log likelihood). The optimization is performed by optimizing each term using stochastic gradient descent on a Unet architecture.

3 Experiments

We use 4500 2D lung images stored on the remote RTX machine. The training process uses an open-source diffusion model [2], which was also used in previous references [3]. We use the parameters listed in Table 1 for the training process and generated 1000 2D lung images from the trained model.

Diffusion steps	4000
Image size	256
Number of channels	96
Number of residual blocks	2
Attention resolutions	1
Learning rate	0.0001
Batch size	128

Table 1: Parameters used in training of the diffusion model from the repository `improved-diffusion` [2].

To confirm that the model does not overfit, that is, the generated images do not result from memorization of the training images in the model, we evaluated the degree of similarity between the generated images and the training images. The metric we use to evaluate the degree of similarity is the minimum L_2 distance between each generated image and the training set after applying the ResNet18 model [4].

For each metric, we split both the training set and the generated set based on aeration, which quantifies percentage of air in the lung image. In terms of the gray-scale image, the aeration corresponds the average normalized pixel value. A lung is considered normal if the aeration is between 60% and 85%. Therefore, we split the training set and generated set into the three categories

$$
\begin{aligned}
&\text{Category 0 :} && \text{aeration between 0\% and 60\%,} \\
&\text{Category 1 :} && \text{aeration between 60\% and 85\%,} \\
&\text{Category 2 :} && \text{aeration between 85\% and 100\%.}
\end{aligned}
$$

Let f be the ResNet18 model and $X^{(i)}_{\text{train/gen}}$ be an image in the training or generated set in category i. The minimum L_2 distance we compute for each image $X^{(i)}_{\text{gen}}$ is

$$
\min_{X^{(i)}_{\text{Train}}} \|f(X^{(i)}_{\text{gen}}) - f(X^{(i)}_{\text{Train}})\|_2, \quad i = 0, 1, 2. \tag{2}
$$

In Fig. 3, we plot the distribution of all the minimum L_2 distances in each of the three categories. The distribution as well as the average minimum distances are found to be similar across all three categories, confirming that our trained model is not biased toward a particular category.

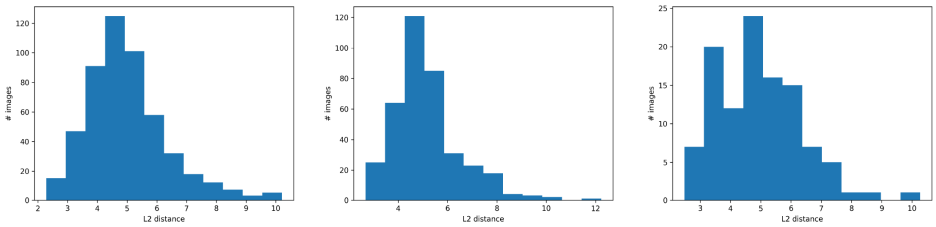

Figure 3: Distribution of minimum L_2 distance between each generated image and the training images under the ResNet18 model. The categories are 0, 1, 2 from left to right. The average as well as the distribution of minimum L_2 distances are found to be similar across all three categories.

In Fig. 4, we show example images where the training image-generated image pair has the minimum distance given in Eq. 2 in each category. We found that even at a distance of 2.2-2.7, the generated image and the corresponding training image are sufficiently different, confirming that our trained model does not overfit the training set but instead generates new 2D lung images.

Category	Training Image	Generated Image	Distance
0			2.280
1			2.675
2			2.4774

Figure 4: Minimum distance training image-generated image pair using the metric in Eq. 2 in each category. The generated image is found to be sufficiently different from the corresponding training image with the minimum distance in each category.

4 Discussion

There are some incomplete aspects of our study. We attempted using the Fréchet Inception Distance (FID) [5] to evaluate the degree of similarity of the training set and the generated set. However, the FID we obtained are exponentially larger than what we would expect, indicating numerical issues in its computation. As the next step after successfully generating images, we attempted performing naive denoising of blurred or impainted original images from the training set, but we were not able to obtain the original images which we anticipated.

Moreover, the original purpose of this study is to use RF data as guidance in the denoising process. Although by the time of this report, we were not able to pursue this direction, this direction is certainly a topic of interest for further investigation.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models, December 2020.

[2] OpenAI. improved-diffusion. `https://github.com/openai/improved-diffusion`, 2021.

[3] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015.

[5] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, January 2018.