Enhancing seismic image resolution using Brownian diffusion bridge model

Bingbing Sun, Saudi Aramco, Abdulmoshen M. Ali, Saudi Aramco, Tariq Alkhalifah, KAUST*

SUMMARY

Seismic images often lack the high resolution needed for proper identification of subsurface structures and potential hydrocarbon reservoirs, and that is due to factors such as limited data acquisition and the attenuation of seismic waves as they travel through the Earth's subsurface layers. Most of the seismic image enhancement algorithms do not account for prior features of high resolution seismic data. Thus, we propose using a generative diffusion model to enhance seismic image resolution. Specifically, we employ the Brownian diffusion bridge model (BBDM) to translate samples from a low-resolution image distribution to one corresponding to a high-resolution image distribution. To address the issue of training solely on synthetic data and improve the generality of the neural network, we adopted a robust training procedure using the know-distillation technique within a "teacher-student" framework. Field datasets demonstrated the robustness and good performance of the proposed method. Additionally, the intrinsic denoising feature of the diffusion model provides an added image-denoising capability for our methodology.

INTRODUCTION

Seismic images play a critical role in understanding the geological composition and resource distribution beneath the Earth's surface. Unfortunately, these images often suffer from limited resolution and significant frequency loss during wavefield propagation. These limitations impact the accuracy of geological interpretation, especially when detecting subtle changes such as lithological facies and subtle geological shifts. Therefore, improving seismic image resolution is essential to enhance the precision of geological analysis and reduce uncertainty in seismic interpretation. Traditionally, methods like seismic deconvolution or sparse spike inversion have been employed to extend the frequency band of seismic images. While effective under favorable signal-to-noise conditions, these conventional approaches (Sacchi, 1997; Chen and Wang, 2018) encounter challenges when dealing with substantial noise. Additionally, achieving enhanced resolution in a structure-oriented manner requires significant effort to maintain lateral coherence.

Artificial intelligence has revolutionized seismic resolution enhancement. Machine learning-based techniques, particularly those using convolutional neural networks (CNNs), aim to predict high-resolution images from low-resolution counterparts (Li et al., 2021; Gao et al., 2023). However, these "image-to-image" predictions often overlook valuable prior information within the images. Furthermore, the training of the neural networks for these methods usually involves synthetic dataset only and this can easily lead to over-fitting, limiting their generalization capabilities to real-world field datasets (Alkhalifah et al., 2022; Zhang et al., 2022).

Recently, the generative diffusion model has demonstrated remarkable performance in realistic image (and even video) generation. Many applications have adopted the diffusion model as a critical infrastructure to enhance performance. Broadly speaking, a diffusion model is a parameterized probabilistic model. It comprises two Markov processes: the forward process and the reverse process. The forward Markov process represents a fixed diffusion mechanism. During this process, Gaussian noise is gradually added to the input data (such as an image) through a series of steps. Conversely, in the reverse Markov process, a parameterized neural network is employed and trained to learn how to reverse this noise-induced transformation and recover the original data. From a probability distribution perspective, the forward process corresponds to the evolution of the complex data distribution toward a simple Gaussian distribution, while the reverse process operates in the opposite direction (Ho et al., 2020).

In our research, we systematically explore the application of diffusion models to enhance seismic image resolution. Specifically, we investigate a type of diffusion model known as the Brownian Bridge Diffusion Model (BBDM:(Li et al., 2023)). Our goal is to enhance images by transforming samples from a data distribution of low-resolution images to a data distribution of high-resolution images. Unlike existing diffusion methods, Brownian diffusion bridge models establish a mapping between the input (low-resolution image) and output (high-resolution image) domains using a Brownian bridge stochastic process. This approach differs from the usual "image-to-image" conditional generation process and contributes to improved model generalization.This work also introduces a novel contribution by utilizing a knowledge distillation technique to address the challenge of generalization for the seismic image resolution enhancement task. Specifically, we demonstrate that by distilling knowledge from a teacher neural network trained solely on synthetic data, the student neural network achieves improved and more robust performance.In subsequent sections, we delve into the fundamental concepts of the generative diffusion model. We then introduce the Brownian Bridge Diffusion Model as our chosen methodology for resolution enhancement. Our training approach leverages knowledge distillation within a "teacher-student" framework. We present compelling results using real-world field datasets, showcasing the efficacy and robustness of our proposed methods.

THEORY

Denoising Diffusion Probabilistic Models

The generative diffusion theory encompasses various approaches, including Denoising Diffusion Probabilistic Models (DDPMs:(Ho et al., 2020)), denoising score matching (Song and Ermon, 2019) and stochastic differential equations (score SDEs:(Song et al., 2020)). While the fundamentals behind these approaches are similar, we opt for DDPMs due to their discrete Bayesian variational formulation, which facilitates comprehensive understanding.

Similar to other diffusion methods, DDPMs involve two essential Markov processes: the forward process and the reverse process. The forward Markov process q is defined as:

$$q(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_\mathrm{T}|\mathbf{x}_0) = \prod_{t=1}^{\mathrm{T}} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad (1)$$

where T is the total number of steps for diffusion, $\mathbf{x}_i, i = 0, ...\mathrm{T}$ represents multidimensional array in different time steps. Specially, $\mathbf{x}_0$ is the sample from the data distribution, while $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_\mathrm{T}$ are the intermediate samples after incremental diffusion (e.g., adding noise to $\mathbf{x}_0$). At each iteration of the forward process, Gaussian noise is added according to:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}), \quad (2)$$

where noise schedule parameter β_t is a scalar value in the range $[0, 1]$ and fixed for each time step and $\mathbf{I}$ is the identity matrix. With proper defined β_t, the forward diffusion process tries to reduce the information content and enhance the noise level. To speed up the training, we can actually sample at arbitrary time step t without referring to the chain formula as follows:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\tilde{\alpha}_t} x_0, (1-\tilde{\alpha}_t)\mathbf{I}). \quad (3)$$

with $\alpha_t = 1 - \beta_t, \quad \tilde{\alpha}_t = \prod_{s=0}^{t} \alpha_s$. With reparameterization, we have:

$$\mathbf{x}_t = \sqrt{\tilde{\alpha}_t}\mathbf{x}_0 + \sqrt{(1-\tilde{\alpha}_t)}\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I}). \quad (4)$$

The schedule of β_t usually results in a distribution $q(\mathbf{x}_t|\mathbf{x}_0)$ to be a standard Gaussian distribution $\mathcal{N}(\mathbf{x}; 0, \mathbf{I})$. By the reverse process, we can generate new data starting from a random signal drawn from $\mathcal{N}(\mathbf{x}; 0, \mathbf{I})$. Specially, the reverse process is structured as a neural network, with parameters denoted by θ:

$$p_\theta(\mathbf{x}_0, \mathbf{x}_1, ...\mathbf{x}_T) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \quad (5)$$

Similar to forward process, transition in the reverse step is defined as Gaussian distribution:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)), \quad (6)$$

where μ_θ and Σ_θ are the mean and variance respectively. In DDPMs, the variance part Σ_θ is usually fixed and not trained.

According to Bayesian variational inference, the loss function for training the DDPMs is given by the Evidence Lower BOund (ELBO). DDPMs reparameterize the mean μ_θ in equation 6 using the added noise ε and simplify the ELBO objective to be a denoising formula :

$$\mathbb{E}_{t, \mathbf{x}_0, \varepsilon} ||\varepsilon_\theta(\mathbf{x}_t = \sqrt{\tilde{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\tilde{\alpha}_t}\varepsilon, t) - \varepsilon||^2. \quad (7)$$

Intuitively, the reverse process can be perceived as making predictions and fitting the introduced noise, denoted as ε. Typically, we employ U-Net for the noise neural network $\varepsilon_\theta(\mathbf{x}_t, t)$, given its superior performance in pixel-level prediction tasks.

The theory described earlier pertains to unconditional generation. However, for typical 'image-to-image' translation tasks (such as enhancing seismic image resolution), we must constrain the generative process based on a given condition, e.g., denoted as $\mathbf{y}$ (representing the low-resolution image). Incorporating this condition in a diffusion model is usually straightforward: we directly inject the condition during the reverse process, leading to the following training objectives

$$\mathbb{E}_{t, (\mathbf{x}_0, \mathbf{y}), \varepsilon} ||\varepsilon_\theta(\mathbf{x}_t = \sqrt{\tilde{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\tilde{\alpha}_t}\varepsilon, \mathbf{y}, t) - \varepsilon||^2. \quad (8)$$

In this approach, the forward process remains intact, implying that the diffusion process is still applied solely to $\mathbf{x}$. The condition $\mathbf{y}$ serves as an additional input only during the reverse process, without any explicit distribution related to $\mathbf{y}$. To elaborate further, consider a non-conditional diffusion model described by Equation 7: during the forward process, we transform the data distribution into a Gaussian distribution and aim to learn the reverse translation (from Gaussian to data distribution). For the conditional diffusion model expressed by Equation 8, our goal remains to reverse from a Gaussian distribution to a data distribution, but with the inclusion of an extra condition. However, in the context of 'image-to-image' translation, we prefer learning to reverse from one data distribution (e.g., the distribution related to the low-resolution image) to another data distribution (e.g., the distribution related to the high-resolution image). To achieve this, we need a slightly different approach, such as the Brownian Bridge Diffusion Model.

Brownian Bridge Diffusion Model

As described, in conventional conditional diffusion model, the forward diffusion process begins with a clean data point $\mathbf{x}_0 \sim p_{\mathrm{data}}(\mathbf{x}_0)$ and evolves toward a standard Gaussian distribution. However, Brownian Bridge Diffusion Model deviates from this trajectory by employing the conditional input $\mathbf{y}$ as the terminal point in a Brownian bridge process. Consider a pair of data points $(\mathbf{y}, \mathbf{x})$, where $\mathbf{y}$ corresponds to low-resolution image and $\mathbf{x}$ corresponds to the associated high-resolution image. The forward diffusion process in the Brownian bridge model follows a Gaussian process and the transitional kernel can be defined as :

$$q(\mathbf{x}_t|\mathbf{x}_0, y) = \mathcal{N}(\mathbf{x}_t; (1-m_t)\mathbf{x}_0 + m_t\mathbf{y}, \sigma_t^2\mathbf{I}), \quad \mathbf{x}_0 = \mathbf{x}, \quad m_t = \frac{t}{T}. \quad (9)$$

$\mathbf{x}_t$ can be equivalently expressed as:

$$\mathbf{x}_t = (1-m_t)\mathbf{x}_0 + m_t\mathbf{y} + \sigma_t\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I}). \quad (10)$$

In 9, T is the total number of steps for the diffusion process, $m_t = t/T$ represents the weight between the source domain image $\mathbf{x}_0$ and target domain image $\mathbf{y}$. The parameter σ_t^2 denotes the variance for the added noise. Following a similar variance-preserving (VP) strategy as in DDPMs, we set it to be

$$\sigma_t^2 = 1 - ((1-m_t)^2 + m_t^2) = 2(m_t - m_t^2). \quad (11)$$

The choice of m_t ensures that the noise tends to zero at both ends while achieving a maximum value of 0.5 in the middle of the Brownian bridge. Notably, from equation 10, we can observe that $\mathbf{x}_0 = \mathbf{x}$ and $\mathbf{x}_T = \mathbf{y}$. This indicates that the forward diffusion process in BBDM bridges the gap between the source and target images. A similar loss function as equation 8 can be formulated to train the BBDM model:

$$\mathbb{E}_{t,(\mathbf{y},\mathbf{x}_0),\varepsilon}||\varepsilon_\theta(\mathbf{x}_t = (1-m_t)\mathbf{x}_0 + m_t\mathbf{y} + \sigma_t\varepsilon, t) - \varepsilon||^2. \quad (12)$$

A naive diffusion model typically requires hundreds of steps for inference. Fortunately, several algorithms have been developed to accelerate the sampling process. The Bayesian Bridge Diffusion Model (BBDM) is particularly versatile, as it can be easily adapted to utilize most of the existing acceleration techniques. For instance, in our current work, we adjusted the model using Denoising Diffusion Implicit Models (DDIM:(Jiaming et al., 2020)), allowing us to achieve accurate inference with just 5 steps. We provide a concise summary of the inference procedures in Algorithm 1.

Algorithm 1 Inference Algorithm by DDIM

Require: total sample steps N_s, sampling times $\{t_i\}_{i=0}^{N_s-1}$, seismic low resolution image $\mathbf{y}$, trained neural network $\varepsilon_\theta(\mathbf{x}_t, t)$
Initial the sampling step $i \leftarrow N_s - 1$
Initial the seismic high resolution image $\mathbf{x}_{t_{N_s-1}} = \mathbf{y}$
while $i > 1$ **do**
Compute $m_{t_i}, m_{t_{i-1}}, \sigma_{t_i}^2, \sigma_{t_{i-1}}^2$
Compute the predicted noise $\varepsilon = \varepsilon_\theta(\mathbf{x}_{t_i}, t_i)$
Compute the high resolution image:
$\mathbf{x}_0 = \frac{1}{1-m_{t_i}}(\mathbf{x}_{t_i} - m_{t_i}\mathbf{y}) - \sigma_{t_i}\varepsilon$
Weight and add noise :
$\mathbf{x}_{t_{i-1}} = (1-m_{t_{i-1}})\mathbf{x}_0 + m_{t_{i-1}}\mathbf{y} + \sigma_{t_{i-1}}\varepsilon$
Iterate to next step $i \leftarrow i - 1$
end while

Knowledge distillation: student beats teacher

In the context of our training, we require a pair of low-resolution and high-resolution images. However, obtaining such a dataset for training in real-world applications is challenging. On the other hand, we can simulate synthetic low-resolution and high-resolution images. Unfortunately, neural network models trained solely on synthetic data struggle to generalize to real-world data due to domain shift. The significant differences in waveform signatures and noise patterns between synthetic and real data pose a challenge. To address this issue, we employ a domain adaptation technique. Specifically, we adopt a teacher-student knowledge distillation approach. Here's how it works: First, using exclusively synthetic data, we train a teacher model. Once trained, we keep the teacher neural network fixed. Next, we feed real low-resolution images to the teacher model and use its output (assumed to be the predicted high-resolution image) as labels for training another student neural network. Importantly, the student neural network only receives the real low-resolution image as input, while the corresponding high-resolution image label is provided by a pretrained teacher neural network. Our experiments demonstrate that this teacher-student knowledge distillation framework works effectively. It tends to produce artifact-free high-resolution images from real low-resolution inputs, making it more robust in practice..

EXAMPLES

In this section, we present two illustrative examples. The first example showcases the effectiveness of the teacher-student knowledge distillation framework when applied to field data. The second example highlights the denoising capabilities inherent in the diffusion model, demonstrating that the learned neural network performs joint denoising and resolution enhancement. We generated 640,000 synthetic samples (as depicted in Figure 1) for training the teacher neural network. The image size for processing is 64 by 64 pixels. In practical applications, we apply an inline section on windowed patches. Both the teacher and student networks utilize the same U-Net architecture, with the student neural network having its channel size halved. Following the recommended knowledge distillation training procedure, we applied the trained model to a field dataset, as shown in Figure 2a. The resolution-enhanced image produced by the teacher neural network (Figure 2b) aggressively boosts high-frequency components but exhibits noticeable artifacts. In contrast, the result from the student neural network (Figure 2c) is smoother and artifact-free. Additionally, in Figure 2d, we analyze the spectra, further confirming that the student neural network mitigates significant noise in the higher frequency range, leading to a more robust outcome. Next, we evaluated the trained student neural network on another seismic image (Figure 3a). The corresponding result (Figure 3b) not only demonstrates improved resolution but also showcases the effectiveness of the proposed diffusion-based resolution enhancement method in denoising, particularly for deeper regions of the model.

CONCLUSION

We introduced the Brownian Bridge Diffusion Model as a powerful tool for enhancing seismic image resolution. Specifically, the BBDM facilitates the transformation of the low-resolution image distributions to their high-resolution counterparts. Our approach involves a robust training procedure that leverages knowledge distillation from a teacher neural network (trained exclusively on synthetic data) to a student neural network. The results for field datasets demonstrate the robustness and excellent performance of our proposed method in resolution enhancement. Additionally, owing to the intrinsic denoising capabilities of the diffusion model, our approach also excels in image denoising.

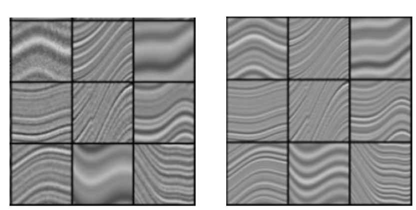
(a) (b)

Figure 1: a) Synthetic low resolution and b) The corresponding high resolution image. Certain amount of noise are added to the low resolution image for data augmentation purposes.

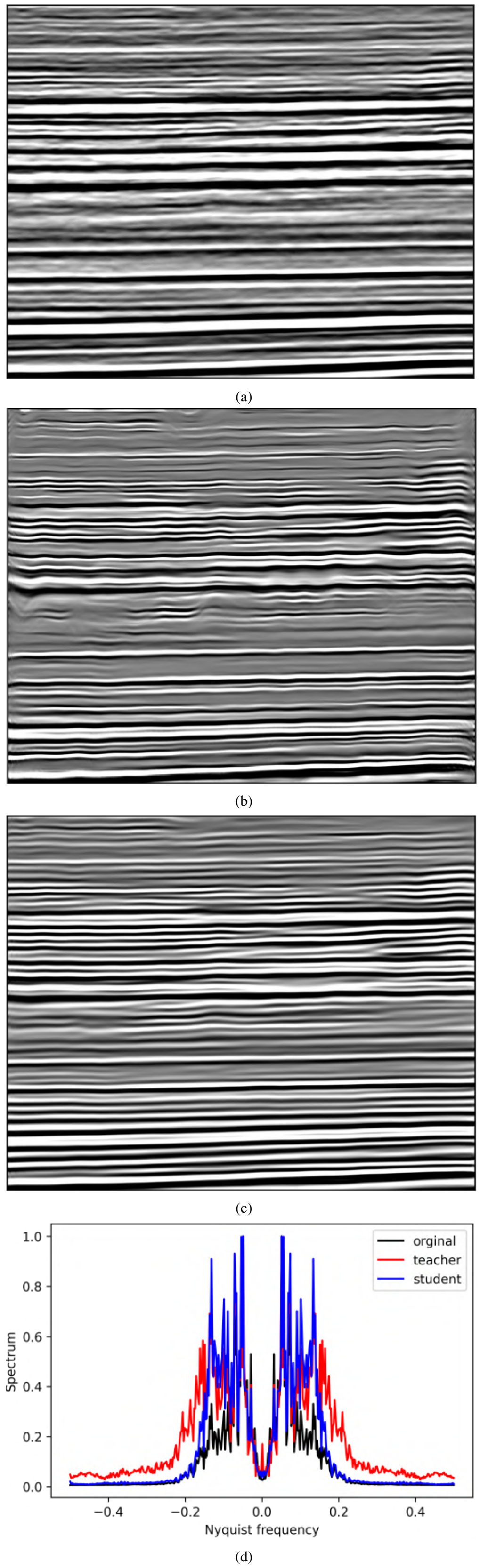

Figure 2: a) Input seismic (low resolution) image; b) Resolution enhanced image by teacher neural network; c) Resolution enhanced image by student neural network; d) A comparasion of their spectra.

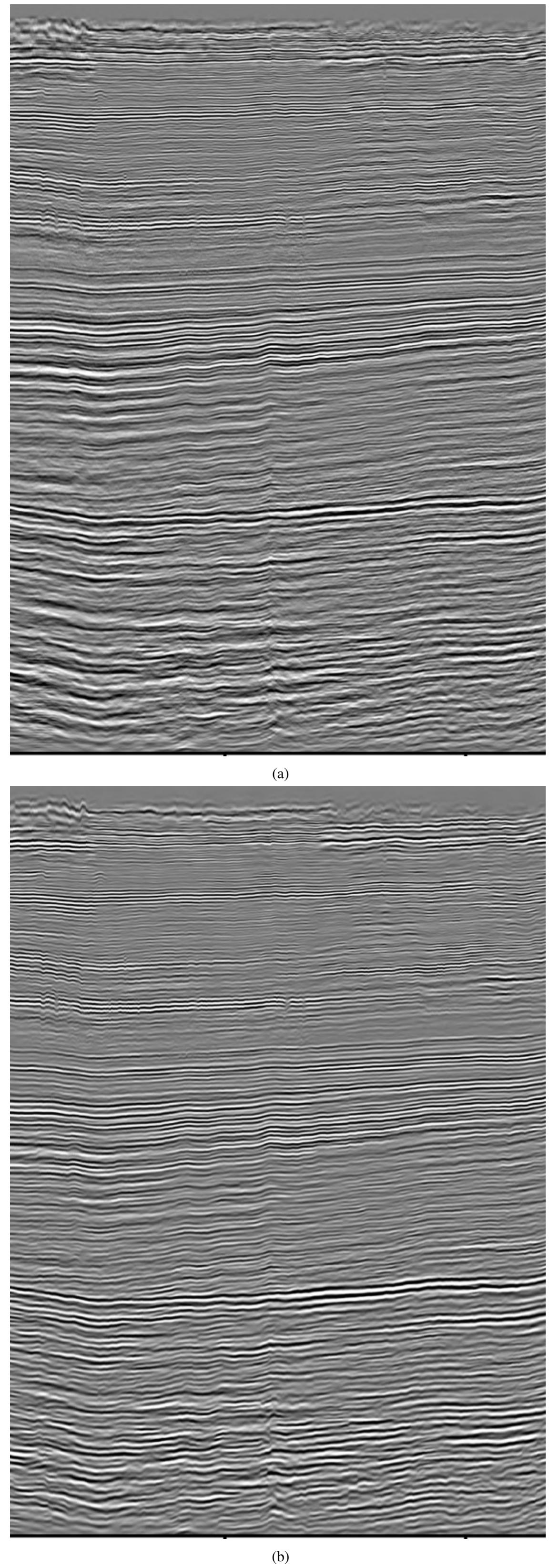

Figure 3: a) Input low resolution image; b) Resolution enhanced image.