Fusion of variable frequency remote sensing images for resolution enhancement

Haidar Al Qatari
Energy Science & Engineering
Stanford University
qatahm0a@stanford.edu

Sohail Waziri
Energy Science & Engineering
Stanford University
swaziri@stanford.edu

Abstract

We investigate the potential of resolution enhancement in subsurface remote sensing images using two different neural network architectures: Standard encoder-decoder, and cycle-GAN. The use of subsurface imaging techniques (seismic data) was a main technological breakthrough in the area of subsurface characterization in the early 1900's. Since then, technology and science kept evolving to enhance what we can see. This is accomplished though frequent data acquisition over large areas. We demonstrate in this paper the potential to limit the size of acquisition and utilize well studied machine learning architectures to enhance older generation subsurface data. We explore this possibility through the use of 3D seismic from Volve field [1], where we synthesized what older generation data would look like by applying a low pass filter. The results indicate an improvement in resolution, although additional investigation is required.

1 Introduction

Geologic characterization of the subsurface is an approach that enables describing the geometry and content of the subsurface. The goal of understanding the geometry and content of the subsurface can include quantifying subsurface storage for climate related endeavors (e.g. CO2 sequestration), characterization of potential hazards and environmental considerations (e.g. underground caves or groundwater mapping), or presence of certain fluids which helps with energy extraction (e.g. oil and gas or geothermal energy). Mapping the subsurface enables scientists and engineers to add important information on a large spatial scale to explain certain phenomena and answer important questions regarding the previously mentioned goals. Mapping subsurface geologic complexities nowadays is generally aided with seismic data. That is achieved by sending sound waves downward into the earth, and then subsequently recording them on the surface. The seismic data is then processed to depict what the layers of the earth may look like in a specific area.

The importance of Remote sensing technology (such as seismic data) lies in its ability to capture more information than what we can observe individually. However, the degree of clarity and detail in one of these images can vary with the technology and method used. We propose fusing overlapping subsurface images of varying frequency content post processing to enhance the resolution and recover features that may not have been recorded due to older technology or more sparse data collection in the widespread legacy subsurface data-sets, which in return, enhances the ability to characterize the subsurface.

Traditionally, enhancements would require professionals who work on sound processing to improve the underlying algorithms/science before turning the recorded sound into an image. Otherwise, data acquisition would either require the use of better technology, and/or denser shot/receiver pairs.

CS230: Deep Learning, Fall 2022, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

However, since we are working with post processed data (i.e. data that has been converted from recorded sound waves to an image depiction of the earth's layers) , the main challenge was finding a large enough data-set for training, in addition to replicating what a real scenario would look like if we decide to synthesize the images. Ideally, our scenario should include one low frequency data-set which attempts to mimic a legacy seismic volume with lower data acquisition specs/technology, and one high frequency seismic volume with higher and more state-of-the-art specs/technology. Due to the proprietary and sensitive nature of such data, we cannot acquire overlapping data-sets of such nature. Therefore, we have taken one seismic volume as ground truth, and have synthesized a lower frequency seismic volume from it to use as our legacy data which we wish to improve.

Two architectures were explored to tackle this problem; an Autoencoder and a CycleGan. The inputs to the models would be a low frequency image/slice from the synthesized seismic volume. The images are matrices of amplitude values representing the contrast between subsurface layers. Therefore, these images are single channel, and can be treated as grayscale images. Once passed through the models, the output would be a higher frequency image which would be compared the ground truth (the initial seismic volume).

2 Dataset

The data-set used in the project is an open source dataset from the Volve field in the North-Sea [1]. The data-set is owned by Equinor, an international energy company based in Norway. The data is a 385 x 605 x 1126 volume of high resolution seismic images. A low resolution dataset is created from the high resolution data by passing the high resolution data through a first order Butterworth filter for the first set. Subsurface layers deposit horizontally in nature and thus the significant variations in the layers are vertical. Keeping the vertical variations in mind, each of the 385 slices are cropped into eight smaller images of size 75 x 350. The last five lines of data in 605 are discarded for divisibility by eight, and the first 450 and the last 320 slices of the data in z-direction are thrown out because data seems to be redundant at those points. To simplify output sizes from convolutions, the images are resized into a squared shape of 128 x 128. Thus, a total of 3080 high resolution and the corresponding low resolution images are obtained. To distribute the data into training/dev/test sets, 70/15/15 proportion is used. A sample high resolution image and the corresponding low resolution images are given in 1

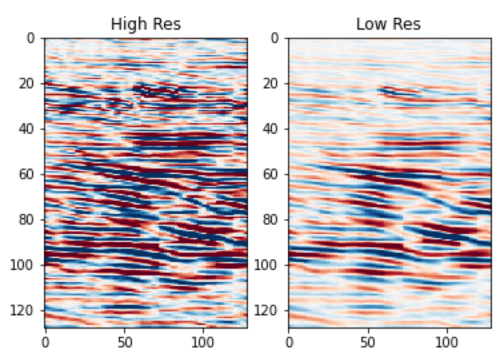


Figure 1: From left to right, high resolution, low resolution (Butterworth filter)

3 Related work

Enhancement of subsurface data is an on-going and extensive area of research. Das et. al. used convolution neural network (CNN) for generating an elastic model of the subsurface using normal incidence seismic data. This showed that it is possible to perform a 1-D inversion of a seismic trace input to a seismic impedance output, which is in essence converting a low frequency signal to a different higher frequency signal. [4]. Li et. al. used deep CNNs to simultaneously enhance seismic image resolution and remove noise, thus allowing detection of detailed stratigraphic features such as faults [6]. Zhang et. al. proposed a two stage framework that utilizes Generative Adversarial

Networks (GAN) to improve resolution of satellite images [7]. Isola et. al demonstrated the use of conditional adversarial networks for image to image translation (such as colorization). Similarly, we can think of our task as translating an image from low frequency to high frequency, and thus restoring features that are likely not present in the lower frequency image [5].

4 Methods

Two different neural network architectures are explored in the context of improving resolution of subsurface seismic images. The architectures are introduced in the following subsections.

4.1 Convolutional Encoder-Decoder

Superresolution of seismic data is investigated using an Encoder-Decoder architecture which involves encoding the low resolution image into a latent space and then decoding from the latent space back to the image dimensions. The resolution of the decoder's output is hoped to be higher than the original input to the encoder. The encoder consists of a series of 2D-convolutional-BatchNormalization layers followed by a flattening layer and a dense layer. The Decoder mimics the operations of the encoder in reverse order using 2D-transpose convolutions. To achieve learning, L2 norm of the difference between high resolution image and output of the decoder is minimized. Since the seismic data involves negative numbers, tanh activation is used. To kick-start code development, Variational Autoencoder architecture from [2] is used and modified.

4.2 Cycle GAN

In this section, a Cyclic General Adversarial Network has been explored for the purpose of image resolution improvement. The model is inspired by the Pix2Pix model utilized by Isola et. al [5]. The generator utilizes a U-net architecture of convolutional layers, deconvolutional layers, and skip connections. Convolutions are followed up by instance normalizations, and leaky ReLu. Deconvolutions are followed up by instance normalization, dropout, and ReLu [3]. Instance normalization is utilized here instead of batch normalization in up-sampling and down-sampling, which normalizes the individual images/slices instead of normalizing the batched data set. This is relevant to this task since each seismic slice may have anomalous amplitude values which significantly benefit from normalizing independently. The first discriminator compares the generated high frequency image to the ground truth. Then, another generator attempts to restore the generated high frequency image to low frequency, and passes that through the second discriminator.

The loss utilized is the sum of losses from the generator, discriminator, and cycle consistency loss [3]. In this case, since we essentially had paired data, random sampling and shuffling of the input data was not enforced to have a direct mapping from input to output.

5 Results and Discussion

5.1 Convolutional Encoder-Decoder

As mentioned in the methods section, a Variational Autoencoder, VAE, was first tested for the superresolution task. Unfortunately, the predicted images were poor as they were converging towards zero even though the ground truth is spread around zero. It is not obvious why the VAE did not work for the purpose of seismic superresolution but one reason could be the small dataset size. Due to time restriction, the VAE was converted to a standard encoder-decoder architecture. Figure 2 compares the output of the encoder-decoder with the input and ground truth images after the model is trained for 10 epochs at a batch size of one and learning rate of $1e^{-3}$. The predicted image does increase the resolution of the image and the run time per epoch is much lower than the cycle GAN (around 20 times faster than cycle GAN) but seems to lack some features such as the dip at position (100, 20). These features are of importance when making informed decisions about potential resources in the subsurface.

In finding a reasonable learning rate. different values (1, $1e^{-3}$, $1e^{-4}$) were tested and losses for the training and dev set were computed. For a batch size of 32, $1e^{-4}$ of learning rate was giving monotonically decreasing loss as opposed to the other mentioned learning rates 3. A better approach

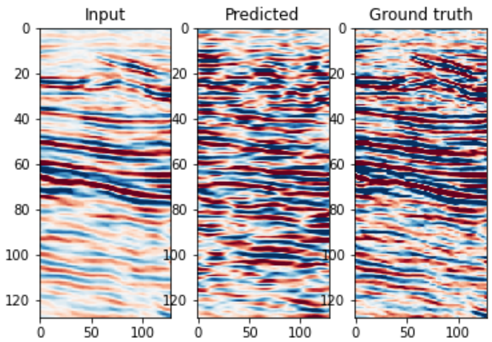
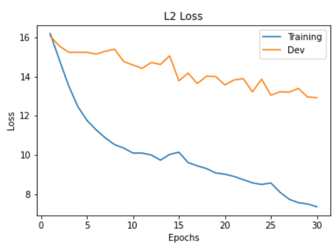
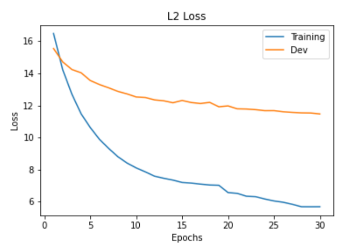


Figure 2: Results on a test set sample after running for 10 epochs at Batch size of one and learning rate of $1e^{-3}$



(a) LR = $1e^{-3}$



(b) LR = $1e^{-4}$

Figure 3: Loss comparison for two different learning rates

of finding the optimum learning rate is to discretize the learning rates on a logarithmic scale and then try to find the optimum one but due to time constraints, other learning rates were not tested.

In addition to the decreasing L2 loss for training and dev sets, the test set was tested based on the Structural Similarity index, SSI. To prevent negative numbers (a necessity by tensorflow ssim), the images were translated by one before feeding into the ssim tensorflow function. Table 1 summarizes the SSI for the different runs. Batch size of 32 and learning rate of $1e^{-4}$ gives the SSI closest to one.

LR, Batch size	SSIM (pre)	SSIM (post)
$1e^{-4}$, 1	0.758	0.864
$1e^{-4}$, 32	0.773	0.890
$1e^{-3}$, 32	0.773	0.862

Table 1: Table to test captions and labels.

5.2 Cycle GAN

Here, a range of minibatch sizes of 40 to 1 have been tested. In addition, a range of learning rates between 2e-4 and 2 have been tested in conjunction with the minibatches. The reasonable values for learning rate and minibatches were found to be 2e-4, and 1 respectively. This was based on systematically iterating these values and measuring the accuracy of the prediction compared to the ground truth using SSIM. Each training iteration was run for 2 Epochs due to time constraints (minibatch=1 runs take around 15 mins per epoch, while minibatch=40 runs take around 5 mins per epoch). The results of running the aforementioned reasonable values can bee seen in 4. The mean SSIM value of comparisson of the training set's low frequency and high frequency images was 0.82. However, The mean SSIM value of the predicted high frequency images and the training set's high frequency images was 0.92. This demonstrates that the predicted images displayed an increase in similarity to the ground truth compared to the inputs after running them through the trained model.

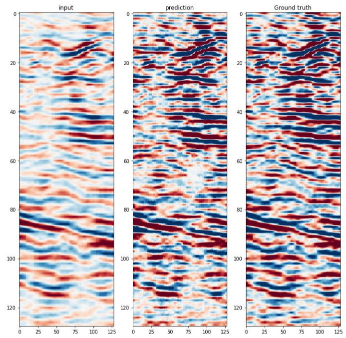


Figure 4: From left to right, low resolution, prediction using Cyclegan and, high resolution (ground truth)

Although the results look promising, adapting the pix2pix Cyclegan architecture to this problem is not direct, as there were additional preprocessing steps that needed to be done to prior to implementation. Image sizing was an issue due to the U-net architecture nature, and convolutional steps were modified since this is not a 3 channel image. In addition, since this data-set contains paired data, all random shuffling needed to be removed. Finally, this data-set contains a large amount of values, and running out of RAM was a recurring issue which we overcame by reducing the data through re-sampling. Although learning progress has been tracked visually, tracking training loss progress quantitatively is done but has not been plotted due to data handling difficulties and time constraint.

6 Conclusion/Future Work

The findings indicate that both a relatively simple encoder-decoder architecture and a more complex cycle-GAN can be used to achieve superresolution in subsurface seismic imaging. However, while Cyclegan takes more time to train, it captures more features of the image than the standard encoder-decoder architecture, which finds a hard time capturing features such as dipping beds. cycle-GAN was trained for low number of epochs, which implies a potential for better results if time and compute power is not a limited resource. Though results show improvement in resolution, it is important to note that the input to the model has been synthesized using a low pass filter, which is a linear filter. We suggest trying non-linear filters, such as median filters, as inputs instead to simulate a realistic situation where there may not necessarily be a linear relationship between two seismic volumes.

7 Contributions

Haidar: cycle GAN code, write-up and slides

Sohail: Encoder-Decoder code, write-up and slides

References

[1] URL: `https://www.equinor.com/energy/volve-data-sharing`.

[2] URL: `https://www.tensorflow.org/tutorials/generative/cvae`.

[3] URL: `https://www.tensorflow.org/tutorials/generative/cyclegan`.

[4] Vishal Das et al. "Convolutional neural network for seismic impedance inversionCNN for seismic impedance inversion". In: *Geophysics* 84.6 (2019), R869–R880.

[5] Phillip Isola et al. "Image-to-image translation with conditional adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.

[6] Jintao Li, Xinming Wu, and Zhanxuan Hu. "Deep learning for simultaneous seismic image super-resolution and denoising". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), pp. 1–11.

[7] Lixian Zhang et al. "Making low-resolution satellite images reborn: a deep learning approach for super-resolution building extraction". In: *Remote Sensing* 13.15 (2021), p. 2872.