# Subsurface Imaging using GANs

**Rustam Akhmadiev**
Department of
Geophysics
Stanford University
Stanford, CA 94305
arustam@stanford.edu

**Rayan S. Kanfar**
Department of
Energy Resource Engineering
Stanford University
Stanford, CA 94305
kanfar@stanford.edu

## Abstract

We explore the potential of using generative adversarial networks (GANs) for subsurface seismic imaging. We pose the problem of correcting the defocused (over- or under-migrated) images of the subsurface. The images used for the training are generated using synthetic dataset through seismic modeling and migration. The performance of the suggested solution is evaluated by comparing the real seismic images with the ones generated using GAN. We expect that the proposed method might potentially allow for fast automated way of building accurate seismic images without good a priori knowledge of the subsurface.

## 1 Introduction

The problem of subsurface imaging using indirect measurements is, in general, a very challenging inverse problem of geophysics, especially for the real-life applications. The complexity of the underlying phenomena, the number of the unknown parameters non-linearly embedded in the governing equations, limitations of the data acquisition and many other factors make this problem highly unstable and ill-posed. At the same time, subsurface imaging is a crucial part of analyzing the geological structures and provides the means for remote studying of the earth's properties.

Probably, the most prominent method of the subsurface imaging hinges upon the surface recordings of the seismic energy (seismic imaging) [Biondi, 2006]. It is based on the analysis of the recorded waveforms that are used for locating objects that diffract and reflect the seismic waves under ground. Being a very powerful tool [Sava and Hill, 2009], seismic imaging, however, requires an accurate knowledge of the subsurface properties, namely the velocities of the seismic waves.

Unfortunately, in the real-life applications neither of the aforementioned components of the successful imaging strategy is readily accessible. Therefore, usually the final image is obtained as a result of an iterative process that includes updating the velocity model and implicitly correcting the subsurface illumination due to the limited and irregular data acquisition. Because of the large amount of data and the huge size of the unknown variables, this procedure, however, is computationally very expensive and requires careful human control in order to produce meaningful result.

In this project we propose using generative adversarial networks (GANs) as an automated tool of building subsurface images from their initial estimates. We believe optimizing over the distance between focused and defocused image's distributions might have greater generalization capacity with respect to other geological scenarios.

## 2 Related Work

Picetti et al. [2018] have demonstrated examples of successful application of GANs (namely pix2pix) using seismic images. The problems that they were trying to solve were, however, limited to compensating for the irregular surface acquisition and subsurface illumination (image inpainting) and increasing the resolution of the seismic images (super-resolution). Our work, on the other hand, is primarily aimed at compensating for the lateral mispositioning of the images with limited a priori information on the velocity model, which, we believe, is a more ambitious and complex problem.

## 3 Dataset

The experiments involve modeling and imaging synthetic datasets available at our disposal [Versteeg, 1994]. Having the true velocity model we model the wave propagation and record the true data. These are used to generate the true image of the subsurface using one of the conventional tools called seismic migration [Stoffa et al., 1990]. We then heavily smooth the true velocity model

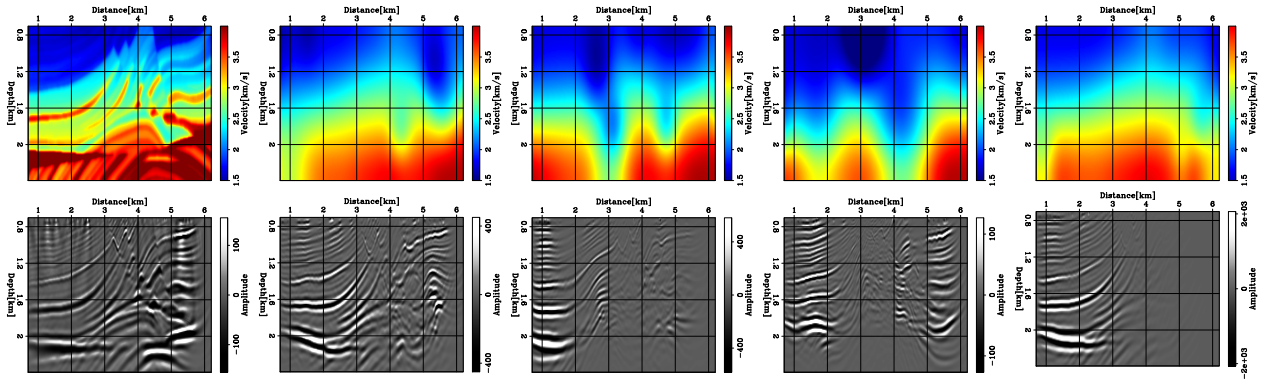



**Figure 1:** The velocity models (top) used for seismic migration and corresponding migrated images (bottom). The leftmost image is the "true focused" image, the rest are the examples of images corresponding to randomly perturbed velocity model.

and perturb it in 100 different ways by adding random Gaussian-shaped anomalies of varying size and amplitudes of 10 - 30% perturbation of the true velocity at a given point. Since the resulting velocity models are not correct, the reflectors in the generated images are mispositioned and the seismic energy gets defocused that leads to the overall poor image quality (Figure 1). It is worth mentioning that these errors in the velocity model are quite large to be corrected for using state-of-the-art physics-based methods for velocity model building. For example, the method of full-waveform inversion [Virieux and Operto, 2009], that is commonly applied nowadays for velocity estimation, is not only computationally expensive, but being a non-convex optimization problem, probably will not be able to converge to the correct velocity model if the starting guess is too far off from the true solution.

In order to generate the training set, we divide every single unfocused image to patches of $141 \times 141$ size along the horizontal axis (decimated to $64 \times 64$ for training purposes). Since the underlying geology is changing somewhat smoothly, the patches coincide with each other with $\sim 60\%$ overlap laterally. Along the depth axis, we use the full extent of the image, because the errors in the velocity model have integral effect (due to the wave propagation), hence, the errors in the upper layers will affect the reflectors at the bottom of the image. To populate the training set with more examples we also flip each patch horizontally. Finally, having 1800 examples of "true focused - defocused" image pair allows us to pose the problem in the framework of deep learning.

## 4 Methods

Generative Adversarial Network [Goodfellow et al., 2014] is a framework based on game theory for likelihood-free training of generative models by optimizing over a learned test statistic. The goal is to train a generator network given a random vector $G(z;\theta)$ to produce samples from the data distribution $p_{\text{data}}(x)$. This is done by introducing a discriminator network $D(x)$, which is trained to distinguish between samples from $p_{\text{data}}(x)$ and $p_{\text{model}}(x)$. The generator tries to fool the discriminator by competing in a zero-sum game. The general objective function for the two-player minimax game is shown in Equation 1.

$$\begin{aligned}\min_{\theta}\max_{\phi} L_{GAN}(G,D) =& \mathbb{E}_{(x,y)\sim p_{data}(x,y)}[\log(D(x,y))]+ \\ & \mathbb{E}_{y\sim p_{\theta}(y)}[\mathbb{E}_{z\sim p_z(z)}[\log(1-D_{\phi}(G_{\theta}(z,y))]]\end{aligned} \tag{1}$$

Training a GAN is notoriously difficult. We start simple at an attempt to understand the effects of the network on the generated images and adjust according to the physical intuition through empirical evaluations.

**Baseline deterministic model (U-Net)**

First of all, in order to develop an intuition about the performance of the generator in GAN, we start off by constructing a U-Net [Ronneberger et al., 2015] for supervised deterministic image-to-image translation. After hyperparameter tuning the final architecture for the encoder-decoder convolutional neural network consists of 7 convolutional and 7 deconvolutional layers with filter sizes starting from $11 \times 11$ and getting smaller with deeper layers to become $3 \times 3$. Spatial striding is used in all layers ranging from 1 to 2 samples. Hyperbolic tangent is used as the activation function for all layers in order to capture both positive and negative samples. This architecture results in the receptive field of the last layer of the encoder being equal to the full size of the input image.

**Pix2pix**

Pix2pix is a supervised conditional GAN framework for image-to-image translation problems [Isola et al., 2017]. The motivation is guiding the generative network with a supervised component consisting of the $L_2$-norm difference between the images. The reason we are investigating this framework is to guide the learning of the generator. We couldn't achieve optimum training in

cGAN only with the cross-entropy objective. The objective function of this model is shown in Equation 2.

$$\min_{\theta} \max_{\phi} V(G, D) = L_{cGAN}(G, D) + \lambda L_{L2}(G) \tag{2}$$

As mentioned above, the generator's architecture follows the one from the baseline deterministic model. The discriminator was chosen by empirical evaluation of the generated results.

### CycleGAN

CycleGAN was proposed by Zhu et al. [2017] for the problem of unpaired image-to-image translation. They tackle this unsupervised problem by using the cycle-consistency principle and having two GANs working together. This allows the information transfer from one domain to another without having an explicit one-to-one image mapping. However, this complicates the training process drastically because of the increased number of hyperparameters.

We have adapted the idea behind the cycleGAN and developed our own version of it. The training is done by minimizing the following objective function:

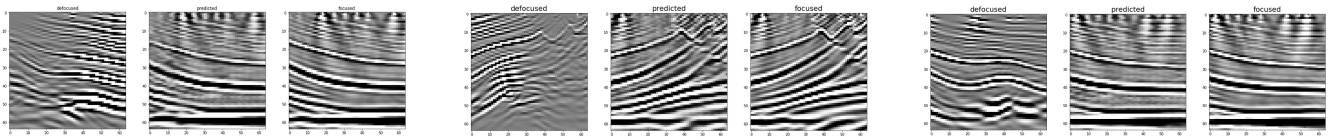
$$V(G, D) = L_{GAN}(D_X, G, X, Y) + L_{GAN}(D_Y, F, Y, X) + L_{cyc}(G, F) + L_{id}(G) + L_{id}(F) \tag{3}$$

The generator of the first GAN ($G$) is learning to map defocused images ($X$) to the focused images ($Y$). While the second GAN's generator ($F$) is trying to learn the inverse mapping from $Y$ to $X$. The discriminators are evaluating whether the generated images from both domains look realistic. We use the CNN encoder as the discriminator that results in the image of $3 \times 3$ size. The cycle-consistency is ensured by the term $L_{cyc}(G, F)$ that is forcing the generated images of each single generator be consistent with the domain of another corresponding generator. The last two terms in the Equation 3 are there to enforce the identity mapping from domain $X$ and domain $Y$ for both generators correspondingly.

## 5 Results

### U-Net

The results from the deterministic U-Net model served as the baseline as explained in the approach section above. The training is performed by minimizing the least-squares problem using Adam as an optimization method, the constant learning rate of 0.001, the batch size of 5 and a small $L_2$-regularization on the weights. The validation set is chosen to be $10\%$ of the defocused-focused image pairs. The result of applying final model to the images in the validation set is shown in the Figure 2 and the convergence history over epochs in the Figure 3. Therefore, in this experiment we have identified the architecture suitable for the problem and hence, this model is used as the generator throughout all the following generative models.




(a) (b) (c)

**Figure 2:** U-Net performance on 3 examples in the validation set: left - defocused, right - focused, middle - predicted image.

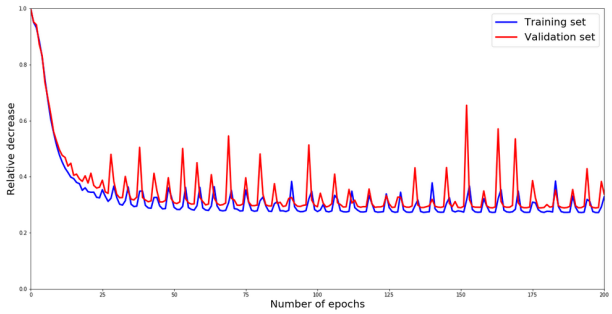
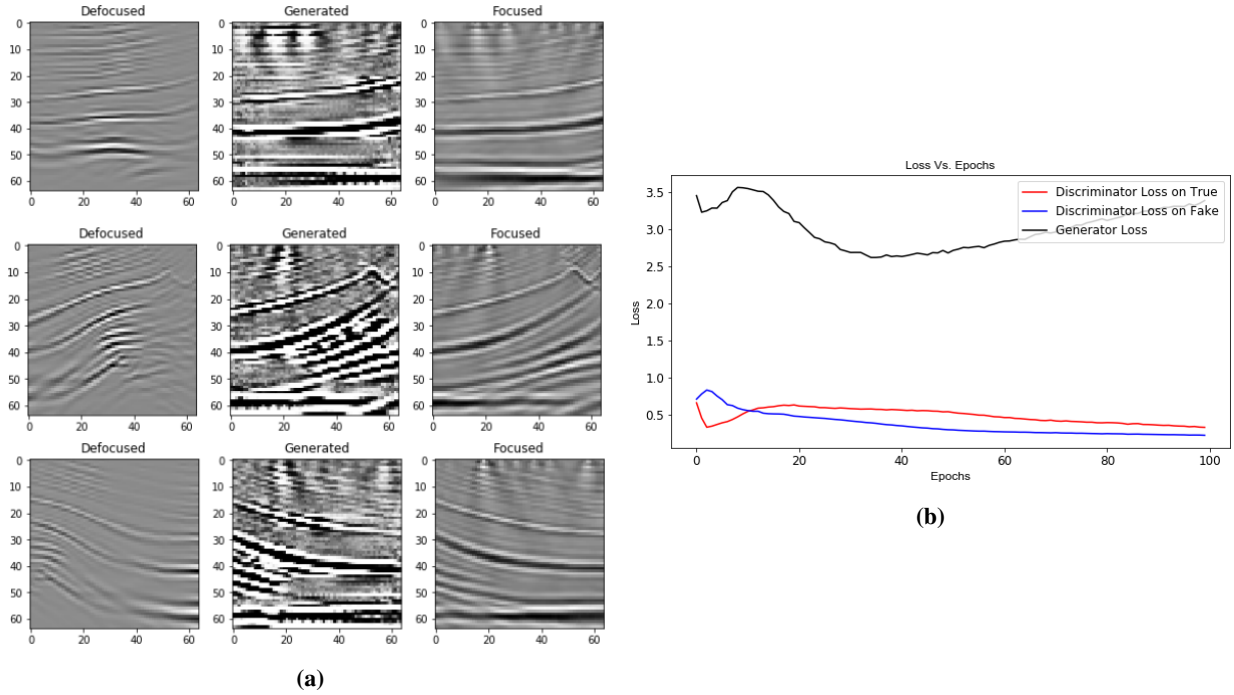



**Figure 3:** Convergence curve over epochs for the U-Net model.

### Pix2pix

The performance of pix2pix model is shown in the Figure 4. The discriminator reaches to its optimum values of 0.5 and the generator is mostly learning over iterations. However, it is still not as good as the baseline model. Although the structure and

geometric changes are captured, the amplitudes are not well generalized. In the exploration industry, the structure, which is the location of the formation is more important than the amplitudes of the image, the values of which bear the information about the rock properties. In this and following (cycleGAN) discriminator we use the CNN encoder with 5 layers and leaky ReLU as an activation function (simple ReLU in the output layer).




**(a)** **(b)**

**Figure 4:** (a) pix2pix generated results (b) pix2pix loss over iterations. We can see that the discriminator lies within the optimum.

## CycleGAN

Following the original idea of cycleGAN, we use $L_2$-norm instead of cross-entropy loss function for both $L_{GAN}$ terms [Mao et al., 2016] in Equation 1. For the cycle-consistency and identity loss we use $L_1$-norm instead, as it has shown to provide higher resolution images throughout the training. We have also found that weighing the terms in the total objective function helps stabilizing the learning process. Therefore, we use a small weight on the discriminator loss in order to slow down its learning and large weight on the cycle-consistency loss for the generators to produce meaningful results from both domains. Throughout the hyperparameter-tuning we observe the importance of equalizing the speed of convergence for all the elements of the cycleGAN. In order to achieve this, we have chosen to use different learning rates for two GANs (0.0001 and 0.00001). The resulting learning curve for 800 epochs is shown in the Figure 5.

Another important modification to the original cycleGAN idea was motivated by the work of Gokaslan et al. [2018] and Yu and Koltun [2015]. We have noticed that using conventional convolutions in the discriminator makes it hard for it to learn the geometric changes in the translated images. Our experiments have shown that this drawback is closely related to the resulting receptive field of the output layer in the discriminator. By using dilated convolutions instead, we were able to increase the receptive field to the full size of the input image without adding extra layers.

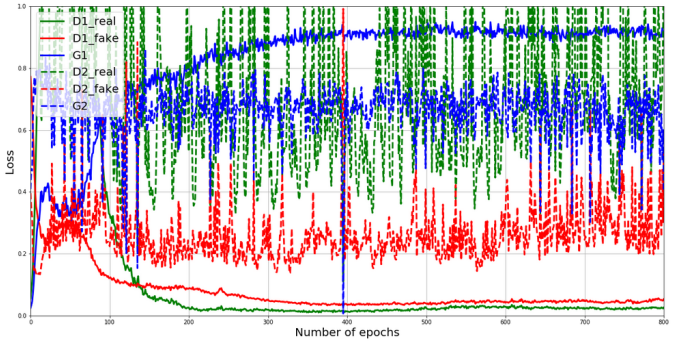



**Figure 5:** Convergence curve for cycleGAN.

The performance of the final cycleGAN model is shown in the Figure 6. The resulting generated images are very close to the ground truth focused image, yet they are not exactly the same. This proves the idea that mild defocusing of the image can be corrected using cycleGAN without having explicit paired examples. The poorer performance (failure cases) is observed when the defocused image is heavily distorted.

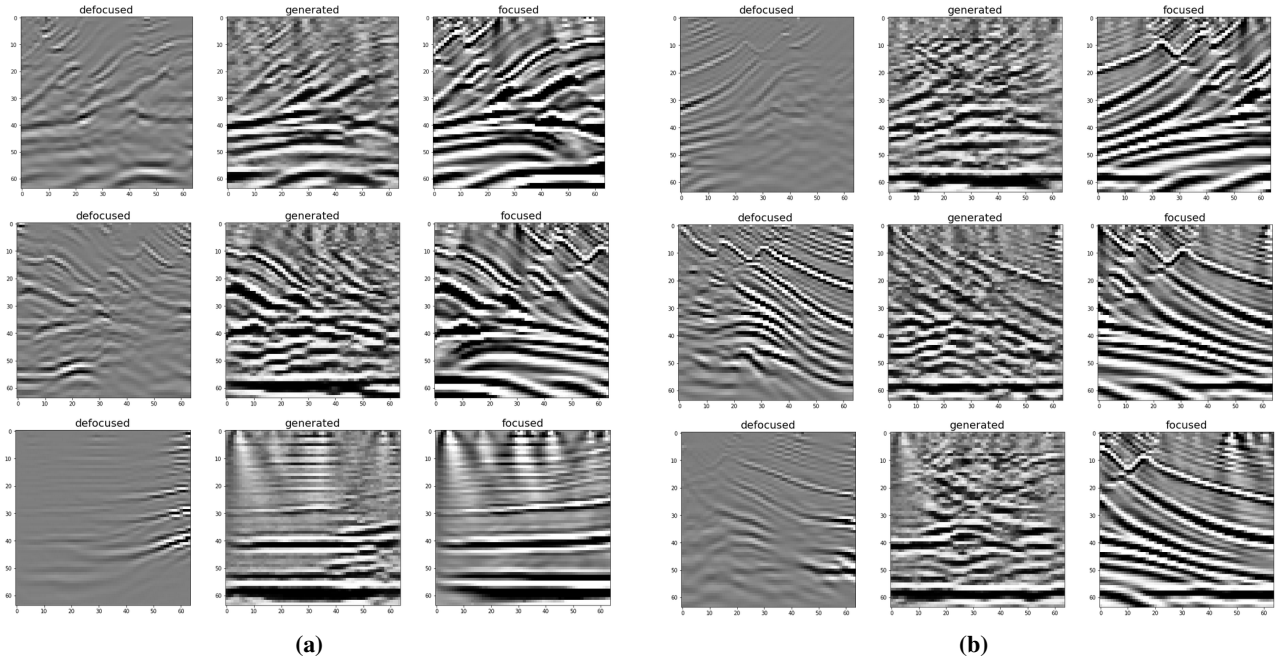



**Figure 6:** Performance of cycleGAN after 800 epochs: (a) - success cases, (b) - failure cases.

## 6 Conclusion

We have shown that deep networks can indeed be used for correcting the seismic image defocusing. We speculate, that the potential benefit of using the generative models in comparison with deterministic is their ability to generalize the transformation over different geological settings. Moreover, having an unsupervised framework (cycleGAN) might be crucial when working with unlabeled and unpaired images, which is often the case in real scenarios. This, however, has yet to be confirmed in the future.

The pix2pix supervised image-to-image translation framework, which guides the cross-entropy objective function with a pixel-wise ground truth loss performs well. The correlation coefficient of the generated images with the focused images are about 0.78 while the defocused images are as low as 0.17. CycleGAN, which is a completely unsupervised framework generates images that resemble closely the focused images. This proves that a defocused image can be translated to focused image without paired supervision. As expected however, both pix2pix and cycleGAN do not perform as well as the deterministic model, although approaching, which is encouraging.

In the future, we plan to experiment with semi-supervised pix2pix model. In this framework, we introduce some pair images at the beginning of training to guide the generator but stop this supervision after a number of iteration. We also want to work on stabilizing the cycleGAN training by testing other objective functions for cycle-consistency such as structural similarity index metric (SSIM) instead of L1. Another point to explore is the Wasserstein GAN cost function which should help analyzing and also stabilize the convergence of the generator and discriminator. Finally, we would like to test the generalization of the models by testing their performance in a completely different geological setting.

## Contribution

Rustam Akhmadiev (enrolled in CS229) has generated and prepared the data for the training using the seismic modeling and migration engines developed throughout his PhD studies. Rustam developed the baseline deterministic model. He also implemented cycleGAN using `pytorch`, performed its hyperparameter tuning and optimizing the results. Rayan Kanfar (enrolled in CS236) implemented and tuned cGAN and pix2pix using `keras`. Code is available at `https://github.com/kanfarrs/Subsurface-Imaging-Using-GANs`

## References


Biondo Biondi. *3D seismic imaging*. Society of Exploration Geophysicists, 2006.

Aaron Gokaslan, Vivek Ramanujan, Daniel Ritchie, Kwang In Kim, and James Tompkin. Improving shape deformation in unsupervised image-to-image translation. In *The European Conference on Computer Vision (ECCV)*, September 2018.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks, 2016.

Francesco Picetti, Vincenzo Lipari, Paolo Bestagini, and Stefano Tubaro. A generative adversarial network for seismic imaging applications. In *SEG Technical Program Expanded Abstracts 2018*, pages 2231–2235. Society of Exploration Geophysicists, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

Paul Sava and Stephen J. Hill. Overview and classification of wavefield seismic imaging methods. *Geophysics*, 28(2):129–256, 2009.

P.L Stoffa, J.T. Fokkernat, R.M. deLunaFreire, and W. P. Kessinger. Split-step Fourier migration. *Geophysics*, 55(4):410–421, 1990.

Roelof Versteeg. The Marmousi experience: Velocity model determination on a synthetic complex data set. *Geophysics*, 13(9): 905–976, 1994.

J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):127–152, 2009.

Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions, 2015.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2017.