Neoscan Image-to-Image Translation using CycleGAN

Gabriel Ronan
University of Michigan
gronan@umich.edu

Abstract

A solution is proposed to mitigate defects in an electromagnetic measurement system and improve the qualitative aspects of the corresponding images. Image-to-image translation using CycleGAN is implemented to find a mapping between defective fast scan images and accurate step scan images. Experiments show that a significant amount of additional training examples may be required to obtain acceptable results. The original CycleGAN code can be found here, the code for this project can be found here, and the dataset can be found here.

1. Introduction

EMAG Technologies is an Ann Arbor-based defense contractor that specializes in RF and electromagnetic solutions. EMAG's Neoscan system is a measurement instrument that uses an electro-optic field probe to simultaneously measure the amplitude and phase of the electromagnetic field of the device under test (an antenna, microprocessor, etc). These measurements are used to identify possible flaws in device design or confirm that the device is working as expected. Neoscan acquires a 2D set of data points in a plane (uniform grid) close to the device. To acquire data points, Neoscan moves the probe across this surface using a raster-like path using a step scan. The probe first sweeps left to right in a straight line across the grid, and then moves down to the next line and sweeps right to left. At every point on the grid, the probe decelerates to a full stop to take a measurement at that point, minimizing structural vibrations. This method takes a significant amount of time to complete if fine spatial resolution is required for a large device.

We developed a new measurement mode called fast scan that does not require an abrupt stop at every measurement point. The probe takes measurements at constant intervals as it sweeps across the device in a continuous motion, only stopping at the edges to change direction. This mode is significantly faster than its predecessor, but in its current state the instrument is not able to perfectly sync the movement of the probe with acquisition of data at the desired points. As a result, the grid of data points may no longer be uniform, and there can be fringe effects on the edges of the resulting image due to an early or late start or stop. An example of a step scan and a particularly poor fast scan are shown in figure 2. The goal is to use a cycle-consistent conditional GAN to map fast scans to step scans. My contributions to the solution are preprossessing of the collected data, designing custom dataloaders, and varying of the GAN hyperparameters within the Tensorflow implementation of CycleGAN.

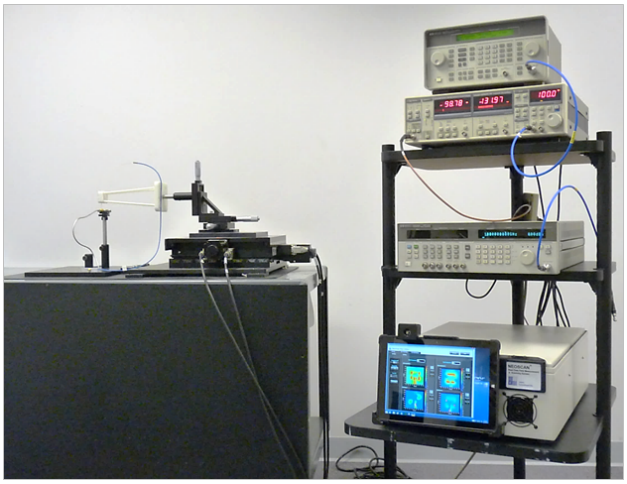

Figure 1. Neoscan system at EMAG Tech (2).

2. Related Work

Generative adversarial networks (GAN) were first proposed by Goodfellow et al. (5) in 2014. A GAN consists of a generator G and a discriminator D. G is a function that maps from a latent space to generate realistic fake data. On the other hand, D is a function that seeks to accurately distinguish the real data from the generated fake data. The process terminates when D can no longer distinguish between the generated fake data and the real data. At this point the generator is now trained to generate realistic images that humans often cannot distinguish from real images. The main

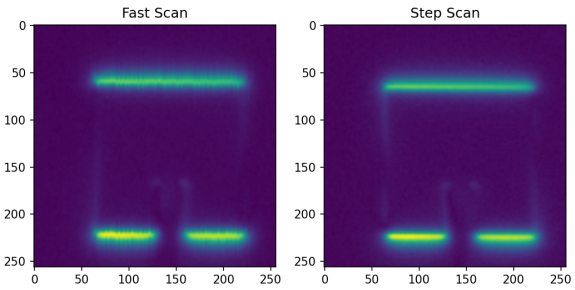

Figure 2. A poor fast scan measurement and an accurate step scan measurement.

advantage of this model is that it can represent very sharp or degenerate distributions, while methods based on Markov chains (3) require the distribution to be blurry. A disadvantage of the vanilla GAN that is shared between its many variants is that *D* must be synchronized well with *G* during training. If *G* is trained too much without updated *D*, *G* may not have enough diversity to model its own distribution over the data.

A natural extension is the conditional generative model (6) $p(\mathbf{x}|\mathbf{c})$, obtained by adding $\mathbf{c}$ to both *G* and *D*. In the case of image-to-image translation that we employ here, the conditional GAN learns a mapping from observed image x and random noise vector z, to y, $G : \{x, z\} \rightarrow y$. The advantage of this model is that it is especially effective on problems where the output is highly detailed or photographic, which is common in image processing and graphics tasks. A disadvantage is that each input image requires a corresponding output image for training. This problem is solved using cycle-consistent adversarial networks (13). This is the unpaired image-to-image translation model that we employ. An additional cycle consistency loss is enforced to further reduce the space of possible mapping functions. This approach is perfect for the implementation of improving Neoscan images; there is no input-output pair relationship for repeated scans of the same device. The advantages of this model include accurate color and texture changes. The disadvantages of this method includes inaccurate geometric changes. This did not seem like a significant problem for this project because the geometries of the scan are nearly identical.

Other variants of image-to-image translation have been explored using both paired and unpaired images. In the paired domain, photographs have been generated from sketches (12), or from attribute and semantic layouts (8). A more recent approach (10) uses a dataset of input-output examples to learn a parametric translation function using CNNs (convolutional neural networks). In the unpaired domain, the CycleGAN has been extended using variational autoencoders with GANs (9). A Bayesian framework has also been proposed (11) that introduces a prior based on a Markov random field from the source image and a likelihood based on multiple style images. Neural style transfer (4) (7) has also shown promise in performing image-to-image translation. The content of one image is combined with the style of another image based on matching the Gram matrix statistics of pre-trained deep features. Many of these models are attractive options to implement in future tests.

3. Methodology

The goal is to learn mapping functions between the two domains X and Y (fast and step) using training samples $x_i \in X$ and $y_i \in Y$. The data distributions are denoted as $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$. The model has two mappings $G : X \rightarrow Y$ and $F : Y \rightarrow X$. D_X and D_Y are adversarial discriminators such that D_X aims to distinguish between images $\{x\}$ and the translated images $\{F(y)\}$, while D_Y aims to discriminate between $\{y\}$ and $\{G(x)\}$. For the mapping function $G : X \rightarrow Y$ and its discriminator D_Y, the objective is:

$$\begin{aligned}\mathcal{L}_{GAN}(G, D_Y, X, Y) =& \mathbb{E}_{y \sim p_{data}(y)}[logD_Y(y)] \\ &+ \mathbb{E}_{x \sim p_{data}(x)}[log(1 - D_Y(G(x))]\end{aligned} \tag{1}$$

where G aims to minimize this objective against an adversary D that tries to maximize it. We use a similar adversarial loss for the mapping function $F : Y \rightarrow X$ and its discriminator D_X: $\mathcal{L}_{GAN}(F, D_X, Y, X)$.

The learned mapping should also be cycle consistent as shown in figure 3 (b); the image translation cycle should bring x back to the original image, $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, a process called forward cycle consistency. Similarly as in figure 3 (c), the system should have backward cycle consistency, $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. This behavior is incentivized by a cycle consistency loss:

$$\begin{aligned}\mathcal{L}_{cyc}(G, F) =& \mathbb{E}_{x \sim p_{data}(x)}[||F(G(x)) - x||_1] \\ &+ \mathbb{E}_{y \sim p_{data}(y)}[||G(F(y)) - y||_1]\end{aligned} \tag{2}$$

Therefore the full objective function is:

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{GAN}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{GAN}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{cyc}(G, F)\end{aligned} \tag{3}$$

To generate the data for this model, each scan image was processed and split into testing and training sets. To artificially increase the size of the dataset, zero mean Gaussian noise was added so that the data would not be biased and we could obtain more samples. The images were upsampled with interpolation and a random shuffle was applied to

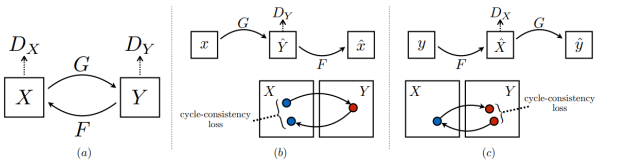

Figure 3. (a) Diagram of the mapping functions G and F, and discriminators D_Y and D_X. (b) Forward cycle consistency loss. (c) Backward cycle consistency loss. (1)

all testing and training sets. The images are then normalized and random jitter and mirroring is applied to the training set to prevent over-fitting. The exact parameters are further detailed in the experiments section.

4. Experiments

A patch antenna was the device chosen for all scans and and electric field amplitude measurements were taken. Therefore every pixel is only a scalar value representing the electric field magnitude. Only 5 fast scans and 5 step scans of the antenna were able to be produced at the time this report was written. The size of each fast scan is 116x112 pixels and the size of each step scan is 119x112 pixels. For each image, 100 artificial images were generated by adding zero mean Gaussian noise. The standard deviation of the noise for each image is equal to the standard deviation of the original image with another random original image of the same type. Using this method we produce 100 fast scan test images, 100 step scan test images, 400 fast scan train images, and 400 step scan train images. This is equivalent to using a single image for the testing sets and the other 4 images for the training sets. A random shuffle was then applied to the training datasets. The images are then up-sampled to 256x256 pixels for faster incorporation into the CycleGAN generator and discriminator models. The normalization and cropping/jitter methods were generalized to work for images of arbitrary size and arbitrary channel number. The number of input and output channels for the generator and discriminator models was changed from 3 to 1 within the imported script *pix2pix.py*. The number of epochs was set to 2. The model began to over-fit when more than 2 epochs were used. This is likely due to the size of the dataset. All other hyperparameters such as batch size and gradient descent parameters were unchanged from the original horses2zebras example.

The baseline is the original fast scan image. If the squared norm of the difference between the predicted image and the step scan image is less than that of the original fast scan image with the step scan image, we can confidently call the project a success. The qualitative shape and jitter of the resulting image will also be taken into account, as that is just as useful when troubleshooting flaws within a device.

The results of training the CycleGAN for 2 epochs are shown in figure 4. There is a clear artifact next to the top bar of the predicted image, and there is a significant amount of noise in the background.

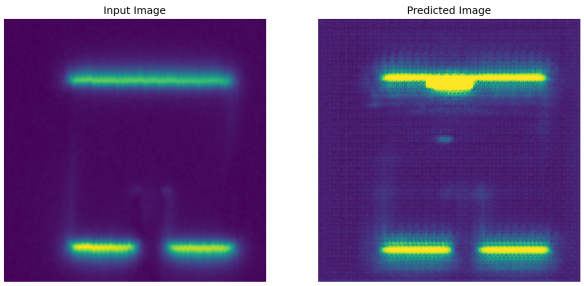

Figure 4. Fast scan test image and generated image after 2 epochs using full dataset.

Unfortunately it is clear that this result is not close to what was desired, and no further statistic analysis is necessary because the image has not been improved. To diagnose the issue, we first turn to the number of data samples. The training dataset consists of 4 original samples of each type. Instead of using all 4 samples, we can reduce the training set to a single sample from which all 400 training images can be generated and observe whether the quality decreases. The result is shown in figure 5.

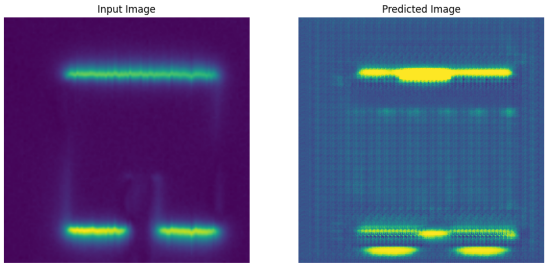

Figure 5. Fast scan test image and generated image after 2 epochs using a single example for the dataset.

Since the quality decreased, it is predicted that the resulting image will improve as more data is added to the training set.

5. Conclusion

The quality of the generated images was not close to the desired result. This is most likely due to an insufficient number of training images as was shown in the experiment. In future work, more data will be continuously

collected. Generating artificial data for training will not be necessary and other methods of image-to-image translation can be tested. At some point, paired images will need to be incorporated to account for a general solution for multiple devices. For example, all devices of the same type should be linked together: patch antenna to patch antenna, microprocessor to microprocessor etc. A hybrid method using semi-supervised data is a potential solution. Although different generators could be trained separately for different devices, it is clear that the defects associated with a fast scan will be shared across all examples, and it will be useful for a single generator to learn this relationship on as many designs as possible.

References

[1] Cyclegan mapping diagrams, https://www.tensorflow.org/tutorials/generative/cyclegan. 3

[2] Neoscan system, emag technologies https://emagtech.com/neoscan-system/#iLightbox[f2ef2de03356a316e6c]/0. 1

[3] Yoshua Bengio, Grégoire Mesnil, Yann Dauphin, and Salah Rifai. Better mixing via deep representations. In *International Conference on Machine Learning*, 2012. 2

[4] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016. 2

[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. 1

[6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. 2

[7] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *CoRR*, abs/1603.08155, 2016. 2

[8] Levent Karacan, Zeynep Akata, Aykut Erdem, and Erkut Erdem. Learning to generate images of outdoor scenes from attributes and semantic layouts. *ArXiv*, abs/1612.00215, 2016. 2

[9] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks, 2017. 2

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015. 2

[11] Rómer Rosales, Kannan Achan, and Brendan J. Frey. Unsupervised image translation. *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 472–478 vol.1, 2003. 2

[12] Patsorn Sangkloy, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Scribbler: Controlling deep image synthesis with sketch and color. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) spotlight presentation*, July 2017. 2

[13] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2017. 2