A Privacy Preserved Image-to-Image Translation Model in MRI: Distributed Learning of WGANs

Berivan Isik, Batu Ozturkler, Tolga Ergen

{berivan0, ozt, ergen}@stanford.edu

Stanford | ENGINEERING
Electrical Engineering

Problem Statement & Motivation

GANs have been widely used in Magnetic Resonance Imaging (MRI) tasks such as image-to-image translation and image reconstruction. However, GANs might suffer from training problems since all GAN architectures demand a large amount of high dimensional samples to be adequately trained [1, 2]. Furthermore, medical data privacy regulations inhibit utilizing patient data in a centralized manner. To mitigate such problems, we implement the distributed learning of WGANs for image-to-image translation tasks.

Our contributions in this work can be summarized as follows:

- Privacy preservation
- Low computational complexity
- Low storage demand

Model-WGAN

- **GAN:** A GAN architecture aims to optimize the following

$$\min_G \max_D \mathbf{E}_{x\sim p_r(x)}[\log(D(x))] + \mathbf{E}_{x\sim p_g(x)}[\log(1-D(x))]$$

p_z: Distribution over noisy inputs
p_g: Generator's distribution over real samples
p_r: Distribution over real samples

- **WGAN:**

$$\min_G \max_D \mathbf{E}_{y\sim p_r(x)}[D(y)] + \mathbf{E}_{x\sim p_g(x)}[D(x)] + \eta \mathbf{E}_{t\sim p_t}[\|\nabla_t D(t)\|_2 - 1]^2.$$

where $t = \alpha x + (1-\alpha)y$ and $0 \leq \alpha \leq 1$.

Model-Distributed Architecture

Given n discriminators, each of them computes an average gradient, $g_k = \nabla F_k(w_t)$, on their local data at the current model w_t and sends these model gradients to the central node where we train the generator. The average of the gradients collected from the discriminators, $\frac{1}{n}\sum_{k=1}^{n} g_k$, is used for the global update as follows [3]

$$w_{t+1} \leftarrow w_t - \eta \frac{1}{n} \sum_{k=1}^{n} g_k$$

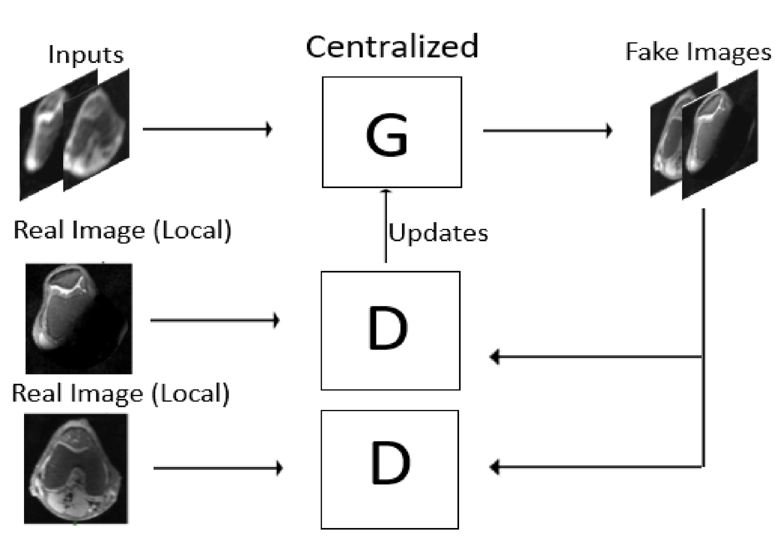

Figure 1: Distributed training of WGANs for MRI image reconstruction.

Datasets & Experimental Setup

Setup: For both G and D, we used a 3-layer Convolutional Neural Network (CNN) architecture with kernel size 5 and stride 2. We then trained each using a first order gradient based optimization algorithm. We implemented the WGAN architecture in a distributed learning setting by utilizing multiple discriminators, and we evaluated the distributed WGAN architecture for the centralized case, for 2 and 5 users.

Data: We used the MNIST dataset [4], and a fully-sampled multi-slice 2D cardiac cine MRI dataset.[5]. The MRI dataset consisted of labeled examples where the labels are fully-sampled, normalized greyscale images, and inputs are undersampled, normalized, greyscale images generated by using variable density undersampling masks. The examples were obtained using a 1.5T MRI scanner where the examples had a matrix size of 180x202.

Results

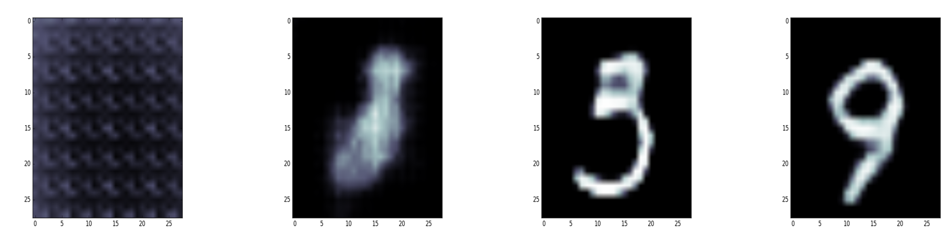

Figure 2: Centralized-MNIST

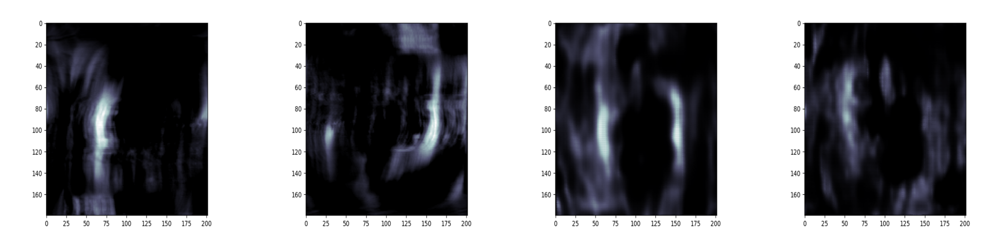

Figure 3: Centralized-the cardiac cine dataset.

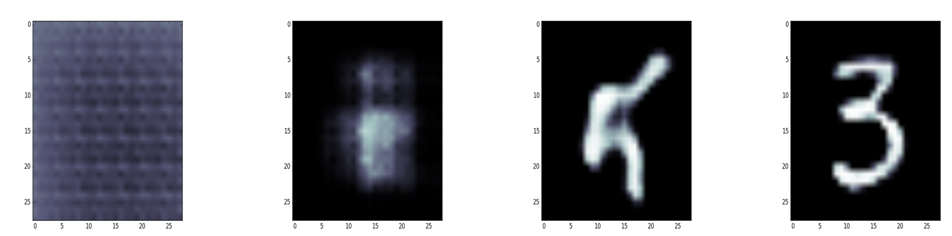

Figure 4: Distributed 2 users-MNIST

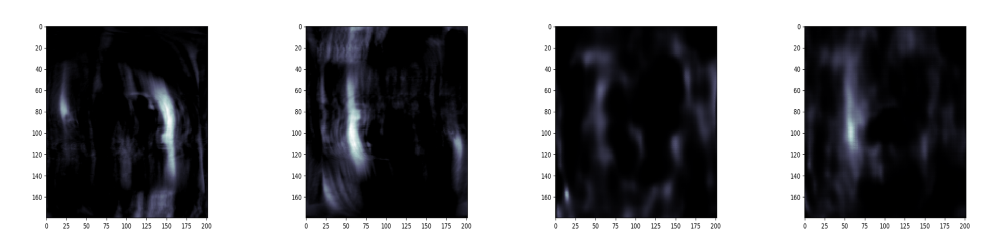

Figure 5: Distributed 2 users-the cardiac cine dataset.

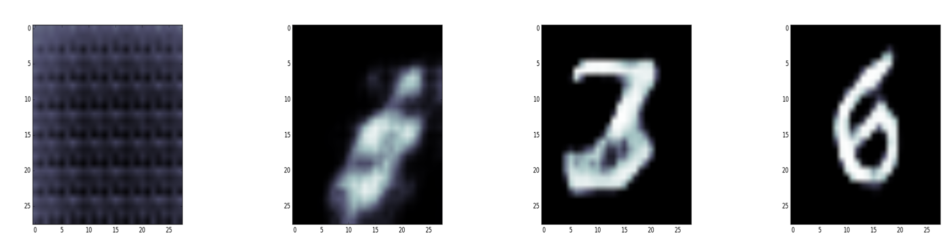

Figure 6: Distributed 5 users-MNIST.

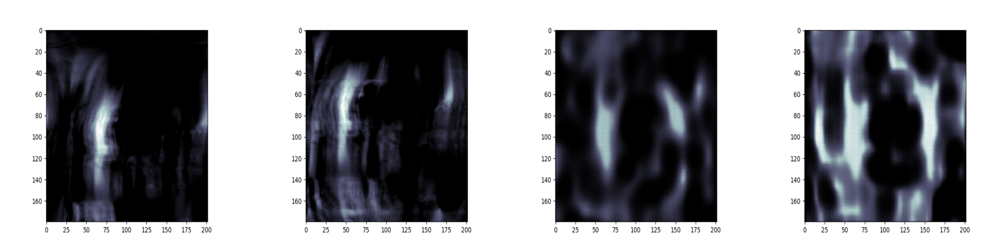

Figure 7: Distributed 5 users-the cardiac cine dataset.

Analysis

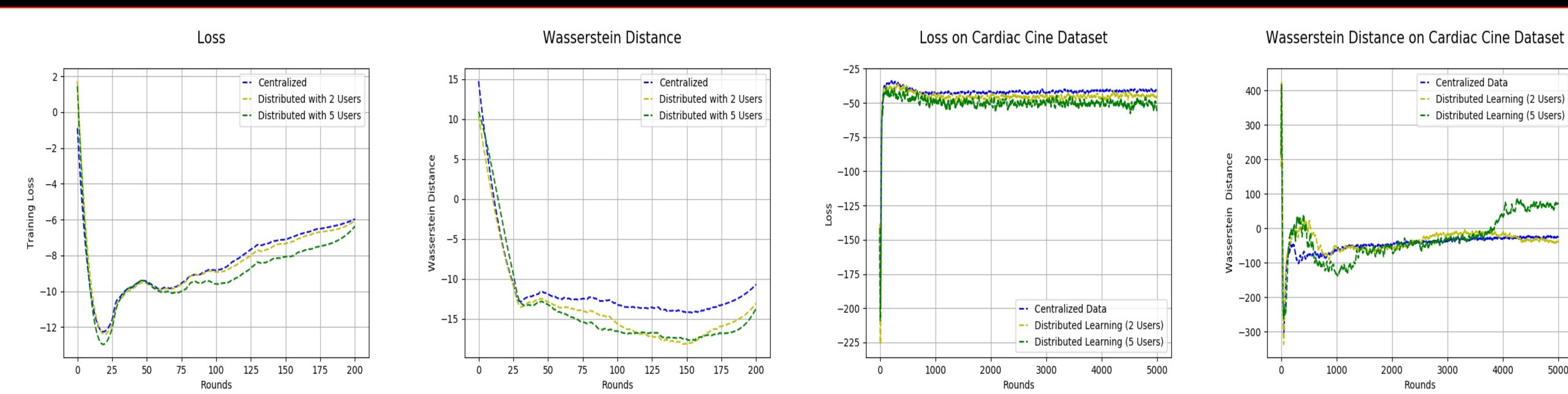

Figure 8: Loss and Wasserstein Distance on MNIST and cardiac cine Dataset.

Table 1: Comparison on MNIST and cardiac cine Dataset.

	Loss-MNIST	Wasserstein Distance-MNIST	Loss-Cardiac	Wasserstein Distance-Cardiac
Centralized	-8.335952	-11.32122	-42.23892	-45.473563
Distributed with 2 Users	-8.401217	-13.544862	-45.66187	-35.75627
Distributed with 5 Users	-8.8937645	-13.643525	-50.28717	-28.68978

Discussion

- We implemented a WGAN architecture in the distributed learning setting and evaluated the performance of our model for MNIST and an MRI dataset.
- Our setting worked succesfully on MNIST dataset as can be seen from the evolution of fake images in Figures 2, 4, 6. Similarly from Figures 3, 5, 7, it is seen that we were able to generate fake images (two left-most images) very similar to the original input images (two right-most images).
- It can be concluded from Figure 8 and Table 1 that distributed learning setting did not cause any increase in the loss or Wasserstein distance which are the basic performance measures for wGANs.
- We preserved each user's privacy since each node stores its local data, which is especially crucial for datasets involving medical information.
- We reduced the complexity and storage demands by distributing both the processing and the data to several nodes, which allows training large scale networks with high dimensional data.

Future Work

- Showing the performance of our model for large-scale tasks, i.e. using larger networks such as ResNet, and evaluating the performance for larger datasets such as ImageNet.
- Compression methods for data transfer between the centralized node and the users can be considered, since communicating updates at each iteration is cumbersome for large neural networks.
- Adoption of different merging techniques to further improve performance, e.g., using a weighted and adaptive average of the gradients.

References

[1] Corentin Hardy, Erwan Le Merrer, and Bruno Sericola. Md-gan: Multi-discriminator generative adversarial networks for distributed datasets, 2018.

[2] Rajagopal. A and Nirmala. V. Federated ai lets a team imagine together: Federated learning of gans. 2019.

[3] H. Brendan McMahan et al. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, 2017.

[4] Yann LeCun. The MNIST database of handwritten digits. `http://yann.lecun.com/exdb/mnist/`.

[5] Christopher M. Sandino, Peng Lai, Shreyas S. Vasanawala, and Joseph Y. Cheng. Accelerating cardiac cine mri beyond compressed sensing using dl-espirit, 2019.