Generative Adversarial Network (GAN) Based Translation Between Medical Image Modalities

Aditi Maheshwari (aditi17@stanford.edu), Ketan Pant (ketan94@stanford.edu) | CS 230, Winter 2019 | Stanford University

**Introduction**

- The objective is to improve the fidelity in which we can translate medical images from one modality to another using Generative Adversarial Networks (GANs).
- Evaluate Unsupervised Image to Image Translation network (UNIT) as baseline and incorporate techniques like Self-Attention, Spectral Normalization to improve results.

**Motivation**

- Unlabeled image data is scarce in medical community because of the high costs of acquisition, rarity of patient conditions and patient confidentiality.
- Translations of medical images between different modalities would reduce the number of times medical devices need to be used and increase the data available to doctors thereby helping the machine learning community utilize disjointed data sets together.

**Dataset**

- Using Dataset provided by the Human Connectome Project.
- Consists of paired T1 and T2-weighted 3D volumes of brain MR images of 1133 patients, split into 900 training and 213 test images in each domain.
- Resolution : 300x256 pixel colour images.

**UNIT**

- Unsupervised Image to Image Translation, or UNIT by Liu et al. [1] is being used to transform MRI images across its T1 and T2 domains.
- In T1-weighted MRI, tissues with high fat content appear bright and compartments filled with water appear dark while it is the opposite in T2-weighted images.
- The model learns the joint distribution between the 2 domains by the "shared latent space assumption".

**Modified UNIT**

- We incorporate techniques such as Self-Attention, Spectral Normalization (SN) and Charbonnier Penalty to improve fidelity of translations.
- Spectral Normalization was shown to improve stability in GAN training.
- Since L2 regularization penalties outperforms heavily, it prevents learning minute details. Hence we use Charbonnier penalty: \( p(x) = \sqrt{x^2 + \epsilon^2} \)

**Self Attention** [2]

- \( f(x) = Wf(x, g(x)) = W_f x, s_{i,j} = f(x_i)g(x_j) \)
- \( h(x_i) = W_h g, \beta_{ij} = \frac{e^{s_{ij}}}{\sum_j e^{s_{ij}}} \)
- \( \theta_i = \sum_j \beta_{ij} h(x_j) \)
- \( \beta_{ij} \) indicates extent to which model attends to \( i^{th} \) location when synthesizing \( j^{th} \) region.

**Architecture**

- Here, \( N=\)Neurons, \( K=\)Kernel size, \( S=\)Stride size.

**Qualitative Results**

- With self attention in encoder, SN in encoder, generator and discriminator, Charbonnier penalty, below is a qualitative comparison of the two models.

**Conclusion and Future Work**

- Even though quantitative results were similar to baseline, qualitatively, our model was better at reproducing intricate details ⇒ search for other evaluation metrics?

**References**
