Capstone Project Proposal
Machine Learning Engineer Nanodegree
Udacity

Raul Pino
August, 2022

Ensembles of Generative Adversarial Networks as a Data Augmentation Technique for Alzheimer research

Domain Background

Data scarcity is a regular problem in research, and in medicine it's especially difficult to find datasets publicly available. The main reason is its rarity, by definition images of anomalies are scattered or not common, and also there are a lot of legal issues that prevent sharing personal information about patients to the rest of the world (Hello Future [1]).

The area I'm interested in is Alzheimer and Dementia research, where there have been some interesting initiatives using Deep Neural Networks for classification (NVIDIA [8] [9]) and GANs for data augmentation (NVIDIA [10], Filippos K. et al. [6]).

Problem Statement

The problem is based on the Kaggle challenge "I'm Something of a Painter Myself" (Kaggle [5]), where the idea is to train GANs, and compare it with the real images to verify how accurate and performant models are. In this capstone project, I'll be training a Deep Convolutional Generative Adversarial Network (Radford et. al [11]), for generating Magnetic Resonance Imaging (MRI) images of brains with Alzheimer and/or dementia, and calling that model CM. Also, I'll train three different models, as results of ensembles of GANs using three different techniques (Yaxing Wang et al. [12]), using the same MRI dataset, and calling those models EM1, EM2, and EM3. The idea is to compare images generated by all models with the real images. CM will function as a control model.

Datasets and Inputs

The chosen dataset used for this capstone proposal it's going to be "Alzheimer's Dataset (4 class of Images) - Images of MRI Segementation" (Kaggle [4]). The data is hand collected from various websites with each and every label verified. Consists of MRIs, having four classes of images, both in training as well as a testing set:

1. Mild Demented
2. Moderate Demented
3. Non Demented
4. Very Mild Demented

This dataset is divided into two folders, training and testing both containing a total of around ~5000 images, each segregated into the severity of Alzheimer and/or dementia, as shown below:

```
Alzheimer_s Dataset
  test
    MildDemented
```

ModerateDemented
NonDemented
VeryMildDemented
train
MildDemented
ModerateDemented
NonDemented
VeryMildDemented

Solution Statement

The solution will consist of training 4 models, one for control, and one per each type of GAN ensemble. The basic idea of a GAN ensemble is that a group or GANs might capture many axis of variations that one individual model wouldn't (Lex Fridman Podcast [2]). In particular, I'll be using the three GAN ensemble designs described by Yaxing Wang et al. [12]: Standard Ensemble of GANs (eGANs), Self-ensemble of GANs (seGANs), and Cascade of GANs (cGANs).

Benchmark Model

For a benchmark model in order to measure the performance of all trained models, I'll be instantiating each model to generate 100 images, giving each folder the name of the corresponding model used. After that, I'll be computing FID (Fréchet Inception Distance) scores (Martin Heusel et al. [3]) using pytorch-fid (Steven Lang [7]), by selecting a random group of real images from the test dataset, and executing FID with each pair or 100 real and 100 generated images, so basically executing pytorch-fid 4 times, one per each model generated folder:

python -m pytorch_fid path/to/dataset_with_real_images path/to/dataset_generated_model_x

After generating these metrics, we can compare and decide which model generated images more similar to the real images.

Evaluation Metrics

The metric used for evaluation will be FID computed for each model in the benchmarking process, to calculate the difference between generated images and real images. In FID, the Inception network is used to extract features from an intermediate layer. Then the process models the data distribution for these features using a multivariate Gaussian distribution with mean μ and covariance Σ. The FID between the real images r and generated images g is computed as:

$FID = ||\mu r - \mu g||\ 2 + Tr(\Sigma r + \Sigma g - 2(\Sigma r \Sigma g)1/2)$

Where Tr sums up all the diagonal elements. FID is calculated by computing the Fréchet distance between two Gaussians fitted to feature representations of the Inception network.

Project Design

Data exploration and visualisation

- Download MRIs dataset.

- List number of images for each class, describe characteristics of images like size and colour, position of brain image (is it centred?), and check if all images share these characteristics and there are no outliers (ie: images with a different size ratio, colour, etc).
- Show a set of images for each class and try to see if there are some big differences at naked eye, and relate this with each class concepts from the anatomical point of view like: different levels of dementia, relation between dementia and Alzheimer, etc.
- In general, the idea is to create some background about Alzheimer and its relationship to the MRIs dataset.

Data preprocessing

- Remove images with no brains or maybe some type of noice different from brain matter.
- Remove outlier images like those with brains positioned different from the centre, or try to centre it.
- Usually MRIs are in black and white, so in case there are images in RGB (red, green, blue dimensions), transform it into black and white images (having 2 dimensions will use less memory while training).
- In case images are different from a square, centre and crop the image to fit it into a square, more focused on the brain matter itself than the black background of MRIs.

Data splitting

- The dataset is already splitted into train and test datasets, so after knowing what is the current ratio between these two, I need to decide if it's useful to change it, maybe moving images from test to train dataset or vice versa could be useful.
- Apply this ratio of train and test dataset and create the final dataset for start training.

Control Model (CM) training

Create a DCGAN, discriminator and generator, from which a generator can be used to create "believable" new MRI brain images, showing a low train loss.

Ensemble Models training

Use the same CM design for the three different models we want to create according to Yaxing Wang et al. [12]:

- Ensemble Model 1 (EM1): Standard Ensemble of GANs (eGANs)
- Ensemble Model 2 (EM2): Self-ensemble of GANs (seGANs),
- Ensemble Model 3 (EM3): Cascade of GANs (cGANs).

For these we also want to create generators creating "believable" new MRI brain images, showing a low train loss for all of them.

Metrics generation: Fréchet Inception Distance (FID)

- Instantiate generators from all models, CM, EM1, EM2, EM3, to generate a set of MRIs of brains.
- Create a folder for each of these models: cm_mris, em1_mris, em2_mris, em3_mris.
- Create a folder real_images with a random selection of real images from the test set folder.
- Use pytorch-fid to generate metrics about FID between the real images and each of the generated group of images:

python -m pytorch_fid path/to/dataset_with_real_images path/to/cm_mris
python -m pytorch_fid path/to/dataset_with_real_images path/to/em1_mris
python -m pytorch_fid path/to/dataset_with_real_images path/to/em2_mris
python -m pytorch_fid path/to/dataset_with_real_images path/to/em3_mris

Metrics visualisation and final thoughts

- Show and relate each of the metrics with some image examples from generated datasets and real images.
- Plot all FID metrics and describe the differences observed.
- Assert if the usage of ensembles, and which type of ensemble, improves accuracy and performance. Try to respond to the question, which model generated more similar images than the rest?.

References

[1] Hello Future (2022) *Generative AI: a new approach to overcome data scarcity*. https://hellofuture.orange.com/en/generative-ai-a-new-approach-to-overcome-data-scarcity/

[2] Ian Good fellow at Lex Fridman Podcast (2019) *Ian Goodfellow: Generative Adversarial Networks (GANs) | Lex Fridman Podcast #19*. https://www.youtube.com/watch?v=Z6rxFNMGdn0&t=3013s

[3] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter (2018) *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. https://arxiv.org/pdf/1706.08500.pdf

[4] Kaggle (2019) *Alzheimer's Dataset (4 class of Images) - Images of MRI Segementation*. https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

[5] Kaggle (2020) *I'm Something of a Painter Myself: Use GANs to create art - will you be the next Monet?*. https://www.kaggle.com/competitions/gan-getting-started/overview/evaluation

[6] Filippos Konidaris, Thanos Tagaris, Maria Sdraka and Andreas Stafylopatis (2019) *Generative Adversarial Networks as an Advanced Data Augmentation Technique for MRI Data*. https://pdfs.semanticscholar.org/56fc/4ac13c24776692e9f130197e7f55ed29408a.pdf

[7] Steven Lang (2021) *FID score for PyTorch*. https://github.com/mseitzer/pytorch-fid

[8] NVIDIA (2018) *Stanford Researchers Develop AI that Can Help Diagnose Alzheimer's Disease*. https://developer.nvidia.com/blog/stanford-researchers-develop-ai-that-can-help-diagnose-alzheimers-disease/

[9] NVIDIA (2019) *AI Study Predicts Alzheimer's Six Years Before Diagnosis*. https://developer.nvidia.com/blog/ai-study-predicts-alzheimers-six-years-before-diagnosis/

[10] NVIDIA (2018) *AI Can Generate Synthetic MRIs to Advance Medical Research*. https://developer.nvidia.com/blog/ai-can-generate-synthetic-mris-to-advance-medical-research/

[11] Alec Radford, Luke Metz, Soumith Chintala (2016) *UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS*. https://arxiv.org/pdf/1511.06434.pdf

[12] Yaxing Wang, Lichao Zhang, Joost van de Weijer (2016) *Ensembles of Generative Adversarial Networks*. https://arxiv.org/pdf/1612.00991.pdf