Least Squared Wasserstein GAN with Gradient Penalty for MRI Tumour Brain generation

Simone Alessia ALESSIA.SIMONE@STUDIUM.UNICT.IT *

1. Introduction

Obtaining Magnetic Resonance Images is a challenging task due to several reasons. Firstly, manual annotation of brain images is a laborious and time-consuming process, often requiring domain expertise and meticulous attention to detail. Consequently, the creation of large-scale labeled datasets becomes prohibitively expensive, limiting the accessibility of reliable training data for robust machine learning models. Secondly, the privacy and ethical concerns associated with sharing sensitive medical data present additional barriers in data collection and sharing. An important research has been conducted this year in Italy from some researchers on the University of Firenze (Valeri F, 2023), building a model able to reconstruct low-contrast images with UNet.

The main objective of this report is to produce a biologically plausible set of tumour and normal brain MRI and, consequently, enhance the stability of a deep classifier to detect different types of tumours.

2. Dataset

The dataset[1] used for this purpose is composed by 3000 labelled images divided into two classes, represented as subfolders. The dataset is balanced as each class contains 1500 images labelled as "yes" which represent tumour brain images and 1500 images labelled as "no" which represent normal brain images. In addition, 60 unlabelled images was provided as test set.

Images are divided into train and validation set with 80-20 percentage, producing 2400 images for the train set and 600 images for the validation set.

Figure 1: (Left) Random normal brain MRI, (Right) Random tumour brain MRI

* Code available at: `https://github.com/alessiasimone/knowledge_discovery_homework_2`

1. https://www.kaggle.com/datasets/abhranta/brain-tumor-detection-mri

3. GAN

3.1. Image Augmentation

To ensure the production of biologically plausible images, image augmentation on this first part is divided between train and validation as follow: images on the train set has been resized to 64 x 64 to ensure the same spatial dimensions for each images and to allow the model to go throw less pixels, then, a random horizontal flip only with a probability of 0.5 has been applied to ensure the correct shape of the brain, then the images has been normalized throw channels with mean and standard deviation of 0.5 to lead the pixels to a range of [-1, 1]. The images belonging to the validation set have been resized to 64 x 64 pixel and normalized as the train set to allows a proper comparison.

Then, models will be fed with batches of 16 augmented images as showed in figure 2.

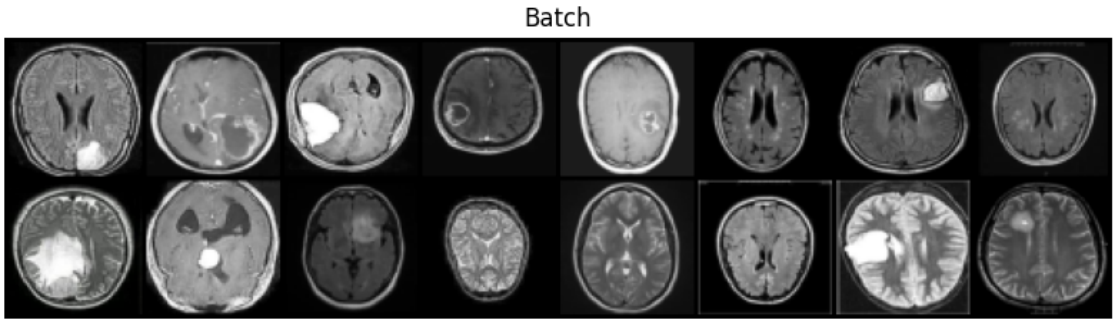

Figure 2: Augmented real batch

3.2. Model Description

For this purpose, two methods has been used together to avoid the most common problems of classical DCGAN architecture (Goodfellow et al., 2014): Wasserstein GAN with gradient penalty was applied to the discriminator to avoid the mode collapse producing a various and diverse images, Least Squared GAN was applied to the generator to ensure a higher quality of images.

Wasserstein GAN (Gulrajani et al., 2017) with Gradient Penalty will penalize the gradient of the discriminator according to a lambda factor equal to 10 to ensure the function to be a Lipschitz function. In addition, the discriminator is trained three times more than the generator to ensure a correct output of synthetic images. This type of GAN makes use of the critic loss instead of binary cross-entropy used in classical DCGAN (which contains a Sigmoid as final layer in addition, compared to the WGAN), which compare the difference between real and fake images. The discriminator, with 2,773,057 parameters, contains an initial embedding layer containing two vectors of size 64 x 64, which will be concatenated to the images to ensure the correct output according to the class. The first convolutional layer accept 3 channels + 1 for the embedding and apply 64 filters of size 4, then, the number of filters increase by a factor of two along the other 4 convolutional blocks. Each block performs a spatial dimension reduction with the stride set to two and contains in addition a

Leaky ReLU as activation function with a negative slope of 0.2 to ensure to maintain some of the negative gradients, and an instance normalization layer.

LSGAN (Mao et al., 2016) is applied to the generator to ensure a higher quality of output images. Here, the MSE loss is applied in order to minimize the difference between the real and fake images. The back-propagation of such a type of loss ensure avoiding vanishing gradient and to reach higher quality of images, which is one of the main cons of WGAN. The generator, with 15,214,595 parameters, contains an initial embedding layer containing two vectors of size 128, which will be concatenated to the noise to ensure the output of image belonging to each class. The first transposed convolutional block accept input with size 256 corresponding to the noise dimension and the embedding dimension and apply 1024 filters of size 4 which decrease throw blocks by a factor of two, until producing an image of shape 3 (D) x 64 (H) x 64 (W). Each block performs a batch normalization to ensure that each batch fit the same distribution and a ReLU as activation function.

Finally, as suggested on the official paper, weights has been initialized with a 0 mean and standard deviation of 0.02.

G optimizer	Adam
learning rate	$2^e - 4$
Beta parameters	0.5 0.99
loss	Mean Squared Error
D optimizer	Adam
learning rate	$1^e - 4$
Beta parameters	0.0 0.9
loss	Critic loss
GP lambda	10
Iterations	3
Z dimention	128

Table 1: Hyperparameters

3.3. Experimental Results

Both models have been trained with an initial number of epochs equal to 200, earlier stopped because of convergence reached.

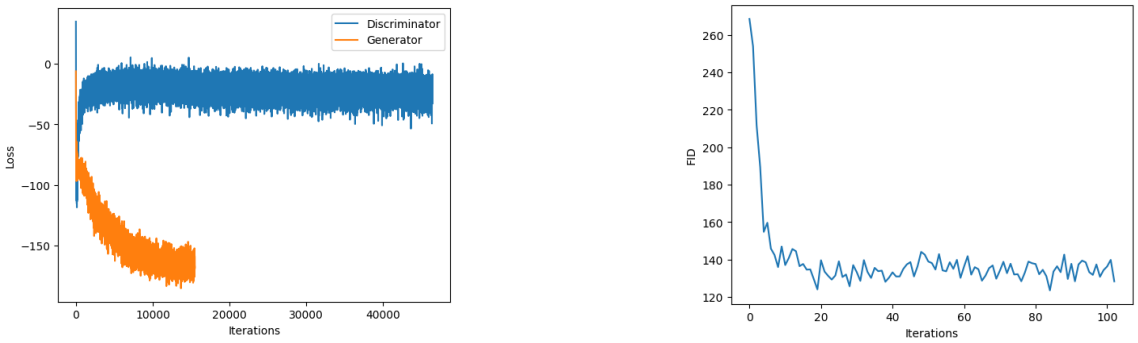

Figure 3: Losses function (right) and FID (left) over epochs

The MSE of the generator decrease over iterations while reducing the difference between real label and fake images, the Critic Loss of the discriminator, trained 3 times more than the generator, increase reaching value around 0 over epochs to reduce the logits between real and fake images. On the other hand, the FID score decrease until reach a convergence around 130 over epochs, comparing the distribution of real and fake distributions on batches of 64 according to the InceptionV3 model.

Figure 4: Comparison between real and synthetic normal brain MRI **FID: 107.62** (left) and real and synthetic tumour brain MRI **FID: 104.22** (right). **Inception score: 1.69 (std: 0.04), Total FID: 106.77**

4. Classification

The classification has been performed on the original dataset and on the augmented dataset in order to understand the impact of the GAN approach. To allow an eclectic usage of the approach, MobileNet V2 (Howard et al., 2017) has been selected as optimal trade-off between efficiency and velocity.

The main reason of such a low latency is the usage of depth-wise separable convolution (Chollet, 2016), first introduced into the deep-learning with the "Xception" network: computations are reduced due to the application of a kernel of size 3 for each channel of the input, followed by a point-wise convolution applied on the concatenated input resulted from the depth-wise kernel. In this way, correlation between pixels and spatial features reduction

are performed on a previous stage and depth is managed by a kernel of size 1 on a lower dimensional feature map.

4.1. Training procedure

For this purpose also, the dataset has been divided into train and validation set with 80-20 percentages, producing 2400 images for the train set and 600 images for the validation set. The test set took into consideration contains 60 images and was used after training for comparison purpose only.

To match the requirement of the pretrained network, images has been resized to 224 x 224 and normalized according to mean and standard deviation of the Imagenet dataset. Then, the training images has been randomly flipped horizontally and vertically and rotated of 15°.

Finally, the model train a batch of 64 images at times, shuffled on the train set by dropping the last batch to avoid odd batches, instead of the validation set.

The classifier was trained for 10 epochs on the original dataset with Adam optimizer, learning rate of $1e-3$ and a weight decay of $1e-5$, and on the augmented dataset with Adam optimizer, learning rate of $1e-4$ and weight decay of $2e-5$.

5. Comparison

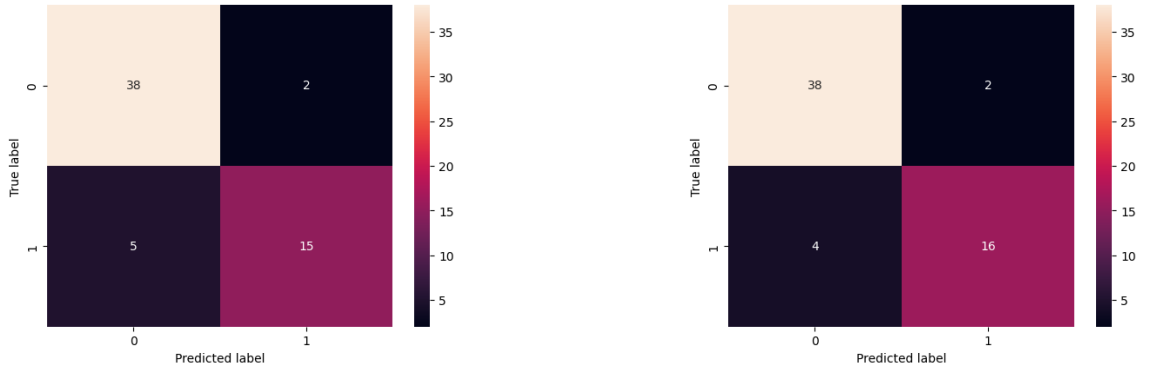

Figure 5: Confusion matrix of MobileNet on original dataset (right) and on augmented dataset (left)

The confusion matrix on figure 5 has been computed with inference on test set. In this field of application, an important metric is the Type-II error, which corresponds to false-negatives observations, such as those positive patients predicted as normal. In this case, Type-II error is at higher risk in both cases compared with Type-I error, however, the impact of GAN allows to slightly reduce this risky behavior.

However, the augmentation with synthetic images produced by LS-WGAN-GP allows to increase the accuracy of **+2%**, with 88.33% on the original set and 90% on the augmented set.

5.1. XAI

Explainable AI is an important and innovative tool which allows to understand the behavior of the algorithm in a visual way. In this case, a Noise Tunneled Guided GradCam has been applied on a positive test image in order to understand what part of the MRI leads the model to predict the tumour. This tool allows to evaluate the activated features of a specific part of the model on a single input, by multiplying the weighted features map by the pixel of the image. In this case, Guided GradCam has been evaluated by taking into consideration the last convolutional block, which is able to produce to the high-level features.

Figure 6: Guided Gradcam of Classifier 1 trained on original images (left) and Classifier2 trained on augmented images (right)

References

François Chollet. Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357, 2016. URL http://arxiv.org/abs/1610.02357.

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

Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville. Improved training of wasserstein gans. *CoRR*, abs/1704.00028, 2017. URL http://arxiv.org/abs/1704.00028.

Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861, 2017. URL http://arxiv.org/abs/1704.04861.

Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, and Zhen Wang. Multi-class generative adversarial networks with the L2 loss function. *CoRR*, abs/1611.04076, 2016. URL http://arxiv.org/abs/1611.04076.

Cantoni E Carpi R Cisbani E Cupparo I Doria S Gori C Grigioni M Lasagni L Marconi A Mazzoni LN Miele V Pradella S Risaliti G Sanguineti V Sona D Vannucchi L Taddeucci A Valeri F, Bartolucci M. Unet and mobilenet cnn-based model observers for ct protocol optimization: comparative performance evaluation by means of phantom ct images. *J Med Imaging (Bellingham)*, 10(Suppl 1):S11904, 2023.