AAEGAN Loss Optimizations Supporting Data Augmentation on Cerebral Organoid Bright-field Images

Clara Brémond Martin[1,2][a], Camille Simon Chane[1][b], Cédric Clouchoux[2][c] and Aymeric Histace[1][d]

[1]*ETIS Laboratory UMR 8051 (CY Cergy Paris Université, ENSEA, CNRS), 6 Avenue du Ponceau, 95000 Cergy, France*

[2]*Witsee, 33 Av. des Champs-Élysées, 75008 Paris, France*

Keywords: Cerebral Organoid, Loss, Adversarial Autoencoder (AAE), Generation, Segmentation, t-SNE.

Abstract: Cerebral Organoids (CO) are brain-like structures that are paving the way to promising alternatives to in vivo models for brain structure analysis. Available microscopic image databases of CO cultures contain only a few tens of images and are not widespread due to their recency. However, developing and comparing reliable analysis methods, be they semi-automatic or learning-based, requires larger datasets with a trusted ground truth. We extend a small database of bright-field CO using an Adversarial Autoencoder(AAEGAN) after comparing various Generative Adversarial Network (GAN) architectures. We test several loss variations, by metric calculations, to overcome the generation of blurry images and to increase the similitude between original and generated images. To observe how the optimization could enrich the input dataset in variability, we perform a dimensional reduction by t-distributed Stochastic Neighbor Embedding (t-SNE). To highlight a potential benefit effect of one of these optimizations we implement a U-Net segmentation task with the newly generated images compared to classical data augmentation strategies. The Perceptual wasserstein loss prove to be an efficient baseline for future investigations of bright-field CO database augmentation in term of quality and similitude. The segmentation is the best perform when training step include images from this generative process. According to the t-SNE representation we have generated high quality images which enrich the input dataset regardless of loss optimization. We are convinced each loss optimization could bring a different information during the generative process that are still yet to be discovered.

1 INTRODUCTION

Cerebral organoids (CO) are brain-like structures that are paving the way to promising alternatives to in vivo models for brain structure analysis. Method implementations such as automatic extraction of shape parameters or size of organoid cultures, requires a large amount of images (Kassis et al., 2019). The scarcity of available data (worsened by the pandemic) is currently a strong limitation to the development of tools to support a more systematic use of CO (Brémond Martin et al., 2021). Data augmentation, a prevalent method in the biomedical domain (Yi et al., 2019), is a possible solution to overcome this issue.

Classical data augmentation strategies transform the input images with a combination of rotations, rescalings, etc, but the content variability that can be observed when acquiring real bright-field images is not reproduced. Deep learning generative methods, called Generative Adversarial Networks (GAN), can solve this problem. Originally introduced by Goodfellow et al, (Goodfellow et al., 2014), GANs are constituted by a generator and a discriminator network trained in an adversarial strategy. Since their introduction GANs have evolved and variations such as CGAN, DCGAN, InfoGAN, Adversarial Auto Encoder (AAE) etc. have been proposed to increase the size of biomedical datasets (Yi et al., 2019).

In this paper, we select and improve the best GAN architecture (AAE) to generate cerebral organoid bright-field images. If the loss effect has already been explored for others biological models in MRI (Lv et al., 2021), to our knowledge, there is no systematic comparative study proposed in the specific context of CO bright-field image generation that gives a quantitative appreciation of this effect. In particular, we

[a] https://orcid.org/0000-0001-5472-9866
[b] https://orcid.org/0000-0002-4833-6190
[c] https://orcid.org/0000-0003-3343-6524
[d] https://orcid.org/0000-0002-3029-4412

Brémond Martin, C., Simon Chane, C., Clouchoux, C. and Histace, A.
AAEGAN Loss Optimizations Supporting Data Augmentation on Cerebral Organoid Bright-field Images.
DOI: 10.5220/0010780000003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 4: VISAPP*, pages 307-314
ISBN: 978-989-758-555-5; ISSN: 2184-4321

are interested in choosing a loss while guaranteeing good quality of the generated data, as well as a good variability of images obtained compared with the inputs in order to improve characterization tasks. The contribution of this paper is to quantitatively investigate the influence of various GAN-based approaches and particularly AAEGAN losses in the specific case of bright-field CO image generation using quantitative metrics from the literature and a dimensional reduction of parameters. The second contribution is to compare data augmentation optimizations using a U-Net-based segmentation task.

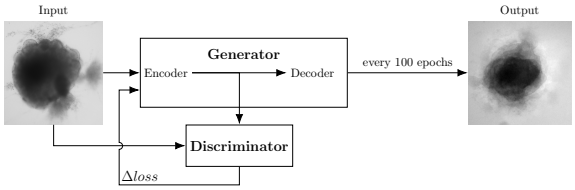


Figure 1: Experimental scheme of AAE supporting data augmentation of cerebral organoids bright-field images. The generator tries to persuade the discriminator that it has generated a true and slightly variable image of input dataset. The discriminator tries to find the true ones. They improve each other by backpropagation, formulated by an objective function based on a loss. Losses variations implemented in this article are symbolized by Δ. Input image is from (Gomez-Giro et al., 2019).

2 METHOD

2.1 Resources

Our dataset is composed of 40 images from an open access database (Gomez-Giro et al., 2019). 20 pathological and 20 healthy CO were numerized with a bright-field microscope over 3 days. The grayscale images are 1088×1388 pixels. However, to compare several networks within a reasonable time, the input images are cropped and resized to 250×250 pixels, maintaining the original proportions.

All algorithms are implemented in Python 3.6 (using an Anaconda framework containing Keras 2.3.1 and Tensorflow 2.1) and run on an Intel Core i7-9850H CPU with 2.60 GHz and a NVIDIA Quadro RTX 3000s GPU device.

2.2 Generative Adversarial Networks

Generative Adversarial Networks (GAN) are made of two competing networks (Goodfellow et al., 2014): the discriminative model (D) computes the probability that a point in the space is an original sample (o) from the dataset distribution(data). However the generative model (G) maps the samples to the data space (z) by an objective function (F). D is trained to maximize the probability of identifying the correct label (true/false) to both generated (g) and original (o) samples. Simultaneously, G is trained to leverage the discriminator function expressed by: $\min_G \max_D F(D,G) = E_o p_{data}[logD_o] + E_g p_g[log(1 - D(G_z))]$.

Various GAN variations have been created since its first implementation. To find the best suited network, we consider five of the most known GAN-architectures to increase the dataset: **GAN** (Goodfellow et al., 2014) is the original implementation; **CGAN** (Yi et al., 2019) gives to the generator input the correct label (physiological or pathological); **DCGAN** (Yi et al., 2019) is constituted by a convolutional neural networks instead of the generator; **INFOGAN** (Yi et al., 2019) uses the generated images at an epoch to train the subsequent; **AAEGAN** (Makhzani et al., 2016) uses an autoencoder as a generator.

During a 1000 epoch duration training step, input images of size 250×250 pixels are used to generate synthetic images. In this work, the original 40 images of the dataset are used to generate 40 synthetic images for a better follow-up by each architecture. The number of images generated are chosen to guarantee no mode collapse, as explained section 3.1.

2.3 Comparative Metrics

We calculate six metrics to compare the quality and similitude of originals and generated images from the various GAN architectures.

FID: Frechet Inception Distance satisfies most of the requirements such as discriminability, or comparisons of efficiency. This metric is used to determine the image quality: a lower FID means smaller distance between generated (g) and input data distribution (o for original). In this equation, μ and Σ are respectively the mean and co-variance of original and generated images: $FID(o,g) = \|\mu_o - \mu_g\|^2 +^T (\Sigma_o + \Sigma_g - 2(\Sigma_o \Sigma_g)^{\frac{1}{2}})$.

SSIM: The Structural Similarity Index compares pixels and their neighborhoods between two images using luminance, contrast and their structure. SSIM has become a standard similarity measure to compare synthetic and natural images even in the biological/medical domain. A high score stands for high similitude: $SSIM(o,g) = \frac{(2\mu_o\mu_g+C1)(2\sigma_o g+C2)}{(\mu_o^2+\mu_g^2+C1)(\sigma_o^2+\sigma_g^2+C2)}$ Constants are added to stabilize the equations.

UQM: We have also implemented the universal quality metric which use the same contrast luminescence

and structures as the $SSIM$. A score of 1 indicate an identical image. This metric is exposed below: $UQM(o,g) = \frac{4\mu_o\mu_g\mu_o g}{(\mu_o^2+\mu_g^2)(\sigma_o^2+\sigma_g^2)}$.

MI: In addition to these well established metrics, we have also calculated the entropy-based Mutual Information between input and generated images in order to measure their correlation (highest score is equal to 1): $MI(O,G) = \sum_{o\in O}\sum_{g\in G} P(o,g)log\frac{P(o,g)}{P(o)P(g)}$.

Blur: To quantify the loss impact on the blurring effect mentioned before, we use the blur metric based on a sharpness quantification of obtained images and local variance value: $\sigma^2 b = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n-1}[p(i,j)-p']^2$ In this equation, m and n are subblocks of images, $p(o,g)$ the predictive residues of images vector and p' its mediane. The lowest score of the global variance corresponds to the sharpest images.

PSNR and MSE: To investigate the quality of generated images, we calculate the peak signal to noise ratio and the mean square error: $PSNR(o,g) = 20\log 10(max(o)) - 20\log 10(MSEo,g)$ and $MSE(o,g) = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(o(m_i,n_j)-g(m_i,n_j))^2$. Here $max(o)$ corresponds to the maximum pixel value of an original image (255).

We calculate FID between each group of generated and the dataset of input images, whereas we compute *Blur* on each image and rendered as a mean. We process $SSIM$, MI, $PSNR$, MSE, UQM between each input and generated images and their mean is rendered for each group.

2.4 Loss Optimizations

To generate the most qualitative and similar images, we choose to optimize the best architecture based upon the previously described metrics. Whatever the architecture considered, all the generated images are somewhat blurry. We choose to overcome this phenomenon by studying how the discriminator loss can influence the quality of the image generation. We introduce six loss with some of them known in literature to resolve similar issues (Kupyn et al., 2018; Lan et al., 2020; Mao et al., 2017).

BCE: The most commonly used loss in GANs is the binary cross entropy (BCE) calculated by (Makhzani et al., 2016) with y the real image tensor and y' the predicted ones: $BCE = -\frac{1}{n}\sum_{i=1}^{n}(y_i*(\log(y_i'))) - ((1-y_i)*(\log(1-y_i')))$. The BCE loss is the baseline of this work. Additionally, we have implemented five discriminator losses which are chosen for their aim to improve the generated images quality with respect to contrast, sharpness, and blur effect.

BCE + L1: First, the original BCE is replaced with BCE and a L1 normalisation (Wargnier-Dauchelle et al., 2019). We hypothesize that such update could improve the quality of the generation as reported in image restoration tasks for instance: $L1 = \frac{1}{n}\sum_{i=1}^{n}|o_i - g_i|_1$ and $BCEL1 = BCE + \alpha * L1$ (10) α is equal to 10^-4, as in the original paper.

LS: In (Mao et al., 2017) the least square loss allowed to avoid gradient vanishing in the learning process step, contributing to create high quality images: $LS = \frac{1}{n}\sum_{i=1}^{n}(o_i - g_i)^2$.

Poisson: In (Wargnier-Dauchelle et al., 2019), a Poisson loss is used to obtain more sensitive results for segmentation tasks: $L_{Poisson} = \frac{1}{n}\sum_{i=1}^{n} g_i - o_i * \log(g_i + \epsilon)$.

Wasserstein and Perceptual Wasserstein: The DeblurGAN was developed to deblur images (Kupyn et al., 2018), using a combination of the Wasserstein and Perceptual loss. Since we are also interested in deblurring the output images, we have tested both losses with the proposed AAEGAN.

We launch loss optimizations on the best architecture during 5000 epochs first to train the model. During the training step, 250×250 pixel input images are used to generate 40 synthetic images every 100 epoch. The global representation explaining this training step on the best architecture is shown Figure 1. We then create a loss value per epoch representation, to highlight when the training has to stop. We have stopped the training at 2000 epochs which corresponds to the plateau before the over-fitting for each loss optimization.

During the testing step based upon the model previously created, 40 images are created by each optimization as explained in section 3.1. We then compare the 40 images generated from each loss optimization based upon the quality and similitude metrics described in section 2.3.

2.5 Dimensional Reduction

The dimensional reduction goal is to observe in the same statistical space if, for each optimization, generated image representations are close or far from the original image representations. We choose to perform a t-distributed Stochastic Neighboor Enbending (t-SNE) dimensional reduction. Contrary to others dimensional reduction methods, t-SNE preserves the local dataset structure by minimizing the divergence between the two distributions with respect to the locations of the points in the map. To avoid subjective or calculated indexes, we perform t-SNE directly on features of images extracted from the GAN networks. t-SNE is constituted with Stochastic Neighbor Embedding where first an asymmetric probability (p) based

on dissimilarities (symmetric) is calculated between each object (x_i), and its probably neighborhood (x_j) (Hinton and Roweis, 2003). The effective number of local neighbors called perplexity (k) is chosen manually: $p_{i,j} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right) / \sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)$. The larger the perplexity, the larger the variance of a Gaussian kernel used to have an uniform induced distribution. Thus we choose the maximal value possible which is 80, the number of individuals in our dataset. To match the original ($p_{i,j}$) and induced distributions ($p'_{i,j}$) in a low dimensional space (the enbending aim), the objective is to minimize the Kullback–Leibler (KL) cost function: $C = \sum_i \sum_j p_{i,j} \log \frac{p_{i,j}}{p'_{i,j}}$. This minimization allows t-SNE to preserve the dataset structure contrary to others dimensional reduction methods (as Principal Component Analysis). Then a Student t-distribution with one degree of freedom is used to avoid the crowding problem (van der Maaten and Hinton, 2008).

We use a momentum term to reduce the number of iterations required (set at 1000 iterations at the beginning) (van der Maaten and Hinton, 2008). The map points have become organized at 450 iterations in a scatterplot. Each point in the map corresponds to the feature vector while the axes are the embedding following the similarity properties i.e. the neighborhood of points. Each run of the t-SNE algorithm generates a different setting of the scatterplot. The points location might be different, but the grouping remains similar. We have launched the t-SNE between original and all the generated features 10 times to validate the similar grouping. We have retrieved the best KL divergence values between original and each generated distributions which could indicate a degree of similitude. A low KL divergence means the two distributions are close.

2.6 Segmentation

To determine the effect of data augmentation with the AAE loss optimizations on a segmentation task, we suggest to consider several training scenarios using a U-Net architecture (Ronneberger et al., 2015). Segmentation allows the extraction of an image content from its background. Various segmentation procedures exist but we have chosen U-Net for its advantages to work well for small training sets with data augmentation strategies, and to have already been used for images of cleared CO (Albanese et al., 2020).

For comparison we perform 40 classical augmentations involving flip-flops, rotations, whitenings, or crops. Second, 40 images generated using an AAE loss optimization are considered. The specific amount of 40 is chosen in order to keep the balance between original images and generated ones in the training dataset, as explained further in section 3.1. To make the performance evaluation more robust, a "leave-one-out" strategy is used, resulting in 40 training sessions (numbers of images in our original dataset). Thus each training is perform on 79 images. We stopped the training at 1000 epochs with an average time of training of more than 1 hour for each leave one out loop (7 cases of augmentations × 40 images = 280 hours almost for the total training step). To compare ground truth cerebral organoid content segmentation (gt) and U-Net (u) ones in various conditions, mean DICE scores are calculated as : $\mathrm{DICE}(gt, u) = \frac{2|gt \cap u|}{|gt| + |u|}$.

To highlight real/false positive/negative segmentation we created a superimposed image composed by the ground truth and a sample of each segmentation resulting from the various trainings. We updated the pixels values in magenta (255, 0, 255) the false positive cerebral organoid segmentations and, in cyan (0, 255, 255) the false negatives.

3 RESULTS

We aim at generating qualitative images of cerebral organoid by GAN strategies to increase the open-source dataset (Gomez-Giro et al., 2019).

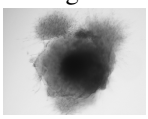
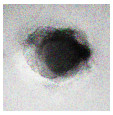
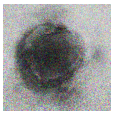
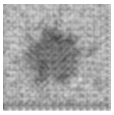
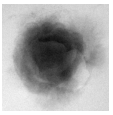
To determine the maximum number of images generated without collapse, we calculate the SSIM between original and generated images. We set that the maximal similitude without creating a twin content is 0.90 (the maximum of similitude between two original images is only of 0.87). Generated images are at the minimum 45 % similar to original images. When we double the generative process, some identical images appear. Thus we choose to generate only 40 images for each case in the testing phase to avoid these duplicates.

3.1 GAN Architecture Choice

To verify the best suited GAN architecture for cerebral organoid bright field images, we first compare the original images and the ones generated using the five architectures.

In Table 1, sample images produced with GAN, CGAN and AAEGAN are the most resembling images compared to the originals. Mode collapse is the most seen in GAN and CGAN architectures as reported in the literature. We can observe also a strong noise for these two architectures results with a white imprint around the shape of the organoid in the GAN

Table 1: Image quality and similitude of cerebral organoids generated by various GAN architectures. Scores within the original range are underlined, best values are displayed in bold.

metric	best	Original 	GAN 	CGAN 	DCGAN 	INFOGAN 	AAE 
FID	low	$0.47 \leq x \leq 0.80$	2.02	2.13	≥ 4	2.89	**1.20**
SSIM	high	$0.65 \leq x \leq 0.71$	0.12	0.12	0.27	0.10	**0.63**
UQM	high	$0.63 \leq x \leq 0.87$	0.79	0.80	0.69	0.66	**0.81**
MI	high	$0.21 \leq x \leq 0.47$	0.17	0.24	0.25	0.19	**0.37**
BLUR	low	$0.10 \leq x \leq 86.28$	2504.24	7561.47	704.48	724.38	**135.93**
PSNR	low	$11.90 \leq x \leq 16.60$	**12.16**	12.64	28.35	28.35	12.89
MSE	low	$93.25 \leq x \leq 106.23$	105.41	103.07	107.12	107.14	**102.72**

case. While AAEGAN generated images are characterized with blurry contours, DCGAN and INFOGAN generate a divergent background making the images difficult to exploit. To verify these observations, we calculate qualitative and similitude metrics, introduced in the section 2.3, by pairing first original images and then original and generated images.

Table 1, presents these results. For the output images, we underline the metric values within range of the original images. We observe only a low proportion of architecture metric within the original range. Indeed, only the AAEGAN and the CGAN answer to only four metrics (UQM, MI, PSNR, MSE) on the seven calculated and only the UQM is within the range of original ones for all the architectures. Regarding FID, SSIM, UQM, and MI scores, AAEGAN generate the most comparable images to the original ones.

In term of quality, this architecture generate the sharpest images, even if the blur index is higher than original images indexes. All the architectures express MSE of between the minimal and maximal values of this metric calculated for the original images. However, regarding the PSNR only the GAN, CGAN and AAEGAN produce images with a score of between the original images limits. To summarize, according to the metric values, AAEGAN is the most suited architecture to generate cerebral organoid images.

3.2 AAEGAN Loss Optimization

Once we have confirmed AAEGAN is the most suited generation architecture for our study, we update the discriminative loss of AAEGAN in order to evaluate the corresponding influence on the generated images quality. Results are shown in Table 2, using the same metrics for quality (PSNR, MSE and Blur) and similitude (FID, UQM, SSIM and MI) of images.

Table 2 shows one of the 40 images generated for each of the six AAE variations. While some of the generated samples are blurry and present a white imprint (BCE, BCE+L1, LS), others show sharper edges and less visible imprints (Poisson, Wasserstein and Perceptual Wasserstein). For this group of losses, only a few of the generated data seem to be identical to the input images: these networks do not suffer from mode collapse.

To quantitatively confirm the visual analysis of the generated images, we calculate comparative metrics between original and generated datasets. Results are shown Table 2. The AAEGAN loss optimizations allow generated images to be within the range on five metrics with the Wasserstein and Perceptual Wasserstein loss (against four for the other loss). Indeed, the Blur index is with these two optimizations within the range of original images. Of the 7 metrics calculated, only FID and SSIM are not in the range for all the optimizations. However, the FID for the Perceptual Wasserstein loss is quite close to the upper bound of the input range (0.82 vs. 0.80).

Quantitatively, Wasserstein and Perceptual Wasserstein networks generate better image quality than other networks, based on lower PSNR, MSE score and Blur index. However they are still higher than the original scores shown in Table 1.

Otherwise, the similarity seems to depend on the architecture. Based upon the FID, the Perceptual Wasserstein loss generates the most similar images. MI is higher for the images generated using Wasserstein networks. However, compared to images produced with BCE and Poisson, images produced with Perceptual Wasserstein loss are not similar to the original dataset regarding the SSIM. Images generated with a Poisson or LS loss have the best UQM index.

The Perceptual Wasserstein loss is the most appropriate loss optimization for generating cerebral organoid images with the AAEGAN. It performs best

Table 2: Sample of images generated for each AAE loss variation. We have calculated metrics on generated images from each AAE loss variations, with the BCE loss as the baseline. Scores within the original range are underlined, best values are displayed in bold.

		Original	BCE	BCE + L1	LS	Poisson	Wass.	Per. + Wass.
metric	best							
FID	low	$0.47 \leq x \leq 0.80$	1.20	1.41	1.33	1.41	1.10	**0.82**
SSIM	high	$0.65 \leq x \leq 0.71$	**0.63**	0.62	0.60	**0.63**	0.62	0.50
UQM	high	$0.63 \leq x \leq 0.87$	0.83	0.83	**0,84**	**0,84**	0.83	0.82
MI	high	$0.21 \leq x \leq 0.47$	0.37	0.39	0.36	0.41	**0.46**	0.42
BLUR	low	$0.10 \leq x \leq 86.28$	135.93	116.30	135.01	106.71	59.84	**59.00**
PSNR	low	$11.90 \leq x \leq 16.60$	13.47	13.74	13.53	13.74	13.17	**12.86**
MSE	low	$93.25 \leq x \leq 106.23$	103.13	103.35	104.01	103.33	103.11	**102.93**

for four metrics and is within the original range for five metrics.

3.3 Dimensional Reduction

To analyze all at once the similitude and the variability of the generated images with AAEGAN loss optimization, we study images in the same statistical space. We implement a dimensional reduction on the features extracted on images in the generative process with t-SNE, see Figure 2. The first observation that can be made is the maintenance of similar positions in the map between original and generated images and that whatever the loss optimization. Some original images constitute a cluster and are almost foreigner to the generated ones. This could be explained by the incapacity of generated images to replicate a background similar to the bright-field acquisition with a light gradient. While at the exterior of the map images generated with a Poisson a Wasserstein or a Perceptual Wasserstein loss are represented, inside the map BCE, BCE+L1, and LS losses are. This observation suggests that each loss optimization could bring different information during the generative process. We compare the KL divergences between original and generated images which remains similar (all results are approximating the null : inferior at 0.3). To summarize, loss optimizations generate similar contents to original images keeping its variability and creating intermediate shapes not seen in original population.

3.4 Segmentation

To illustrate the influence of generated images by an optimised AAEGAN against classical data augmentation, we suggest to tackle a segmentation task in a leave-one-out strategy (n=79 for training and n=1 for testing). We choose the classic U-Net architecture and

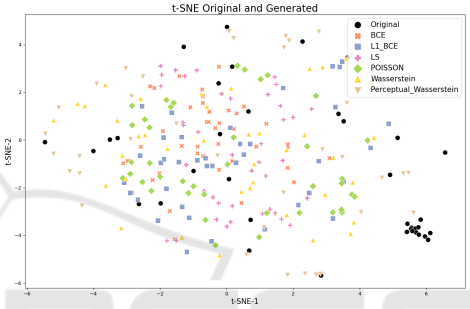


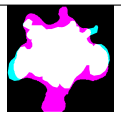
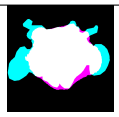
Figure 2: t-SNE representation of original and generated images with optimized AAEGAN.

we consider the different losses to compare the segmentation performance for each data augmentation.

Psychovisually, samples in Table 3 are the best segmented with a training involving images resulting from the AAEGAN Perceptual Wasserstein optimization. They show the less false positive and negative segmentations compared to others AAEGAN optimizations and to classical data augmentation. Quantitatively the mean DICE index highlight the segmentation performance. Results are summarized in Table 3. Mean DICE index is higher for segmented images with Perceptual Wasserstein augmentation, in accordance with the selected visual illustration.

In conclusion, images generated from Perceptual Wasserstein AAEGAN allow a more accurate segmentation than other AAEGAN loss, in accordance with previous results on quality. The influence of the Perceptual loss combined with Wasserstein distance, such as a data attachment term based on the difference of generated and images features maps, improve their sharpness and textural information, making it a viable strategy for data augmentation in this context.

Table 3: Sample Cerebral Organoid image (left) with ground truth (GT) segmentation (our baseline), compared to Classical and AAEGAN-based data augmentation, and the corresponding Mean DICE index and Standard Deviation (SD). Pixels are colored according to the following legend: Black and white represent respectively true negatives and true positives while magenta highligths false positives and cyan a false negatives. The best Dice index is displayed in bold.

GT	Classic	BCE	BCE + L1	LS	Poisson	Wass.	Per. + Wass.
							
Mean DICE	0.87	0.85	0.87	0.86	0.87	0.88	**0.90**
SD	0.03	0.05	0.05	0.14	0.09	0.05	0.04

4 DISCUSSION

This paper presents for the first time to our knowledge data augmentation of cerebral organoids bright-field images, using an Adversarial Autoencoder with various loss optimizations. The Perceptual Wasserstein discriminator loss optimization outperforms the state-of-the-art original article to generate these images, according to the metrics used. Nevertheless, our dimensional reduction experiments suggest all the loss optimizations could bring some variability to the generative process while remaining similar to the original dataset. Approaching a segmentation experiment, images generated with a Perceptual Wasserstein loss could bring a better precision to a segmentation task. Other losses may be interesting for different tasks.

Synthetic images generated with AAEGAN are coherent with original dataset quality contrary to other architectures, with almost no collapse mode but also adding a sought diversity similar to the acquisition of the original dataset. An update of this architecture (for example replacing the autoencoder part by a U-Net) may improve these results. Results remains exploratory with the mentioned small dataset we used. Improvements could be augmenting the number of input images for all of the architectures, increasing the training time for DCGAN and, giving to the INFOGAN generator only high-quality images to avoid the divergence behavior. We only optimize the best architecture for time consideration, but the effect of loss variations on others architectures may be interesting to quantify.

The Perceptual Wasserstein loss optimization of AAEGAN performs best according to metrics. Other loss optimizations show also high similitude, though with a lower quality. However, the dimensional reduction experiment suggests that several loss could be used to generate more images and a good diversity enriching the original training set. In this context, we plan to explore what type of information each loss brings during the image generation. We aim at trying others embedded losses (already used for segmentation tasks) during the generative process based upon high level prior like object shape, size topology or inter-regions constraints (El Jurdi et al., 2021). These losses could be used on condition that the morphological development of CO is better characterized.

Attempting to distinguish the contribution of each loss optimization, this strategy can potentially bring better pixel-wise precision for segmentation tasks. Shown here as a proof of concept, using a U-Net architecture, we demonstrate the Perceptual Wasserstein loss can fruitfully enrich the original dataset. This may also show a kind of regularization achieved by the Perceptual loss leading to a good variability of generated data without being too generic. The contribution of others loss could not been highlighted in this task. Nevertheless, segmentation could be even more appropriate with algorithms suited for small datasets or increasing the training step.

We plan also to train the segmentation task with all the generated images (whatever the optimization) in order to observe the modulation of its accuracy. Indeed, we plan to extract morphological parameters, such as areas, perimeters or higher-order statistics needed for the growth follow-up of cerebral organoid cultures on segmented images. In this work, we only segmented organoid vs. non organoid regions. We aim at reproducing the same work differentiating the peripheral and the core zones of the cerebral organoid in these images.

There is still room for improvement in the proposed AAEGAN network strategy. First, to propose a quantitative evaluation of the generalization of the results obtained on cerebral organoid bright-field images : we would like to use this methodology on others bright-field biomedical images. Biological experts aim at psychovisually evaluating the generated images, and strengthen the quantitative evaluation proposed. In particular, we project to validate the suitability of the metrics we use and observe the training effect on the segmentation task with only validated images by biological experts.

Second, we chose to focus only on the use of an AAEGAN architecture to generate our images, we aim at comparing these results with other types of GANs such as CycleGAN or PixtoPix using U-Net in its generator (Yi et al., 2019). Then, we have to mention the resembling of original images in the t-SNE right corner. It appears generated images could not really generate a similar background such as the lightning gradient of some bright-field acquisition in white light. To resolve this issue, we aim at studying the effect of a similar bright-field background injection during the generative process.

Finally, given the input dataset containing physiological and pathological models of CO, it would be interesting to investigate the generation of specific pathological content in future studies.

5 CONCLUSION

This study answer to the first emerging issue in the cerebral organoid field highlighted in (Brémond Martin et al., 2021) i.e the lack of datasets. These first results show that small databases augmentation of cerebral organoids bright-field images is possible using GANs. Particularly the AAEGAN Perceptual Wasserstein loss optimisation generates the most qualitative content, remains similar to the original dataset and images it generates are useful to improve a segmentation task. However it remains to discover what kind of information other loss optimizations with coherent diversity to the initial dataset could bring during the generative process. This data generation strategy will be valuable to develop characterization methods on CO by enabling large statistical study, but also to develop deep-based approaches for classification and characterization of the various morphologies. Such characterization could help to better understand the growing process once in adequate cultures and help to use cerebral organoids as models for neuropathological disease or for testing therapeutics.

REFERENCES

Albanese, A., Swaney, J. M., Yun, D. H., Evans, N. B., Antonucci, J. M., Velasco, S., Sohn, C. H., Arlotta, P., Gehrke, L., and Chung, K. (2020). Multiscale 3D phenotyping of human cerebral organoids. *Scientific Reports*, 10(1):21487.

Brémond Martin, C., Simon Chane, C., Clouchoux, C., and Histace, A. (2021). Recent Trends and Perspectives in Cerebral Organoids Imaging and Analysis. *Frontiers in Neuroscience*, 15:629067.

El Jurdi, R., Petitjean, C., Honeine, P., Cheplygina, V., and Abdallah, F. (2021). High-level prior-based loss functions for medical image segmentation: A survey. *Computer Vision and Image Understanding*, 210:103248.

Gomez-Giro, G., Arias-Fuenzalida, J., Jarazo, J., Zeuschner, D., Ali, M., Possemis, N., Bolognin, S., Halder, R., Jäger, C., Kuper, W. F. E., van Hasselt, P. M., Zaehres, H., del Sol, A., van der Putten, H., Schöler, H. R., and Schwamborn, J. C. (2019). Synapse alterations precede neuronal damage and storage pathology in a human cerebral organoid model of CLN3-juvenile neuronal ceroid lipofuscinosis. *Acta Neuropathologica Communications*, 7(1).

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. *In Proceedings of NIPS*, page 2672–2680.

Hinton, G. and Roweis, S. (2003). Stochastic neighbor embedding. *Advances in neural information processing systems*, 0(no):857–864.

Kassis, T., Hernandez-Gordillo, V., Langer, R., and Griffith, L. G. (2019). OrgaQuant: Human Intestinal Organoid Localization and Quantification Using Deep Convolutional Neural Networks. *Scientific Reports*, 9(1).

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., and Matas, J. (2018). DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8183–8192.

Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y., and Zhou, X. (2020). Generative Adversarial Networks and Its Applications in Biomedical Informatics. *Frontiers in Public Health*, 8.

Lv, J., Zhu, J., and Yang, G. (2021). Which GAN? A comparative study of generative adversarial network-based fast MRI reconstruction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2200):20200203.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B. (2016). Adversarial Autoencoders. *International Conference on Learning Representations*.

Mao, X., Li, Q., Xie, H., Lau, R. Y. K., Wang, Z., and Smolley, S. P. (2017). Least Squares Generative Adversarial Networks. *IEEE International Conference on Computer Vision (ICCV)*, pages 2813–2821.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*. arXiv: 1505.04597.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using tsne. *Journal of Machine Learning Research*, 9(1):2579–2605.

Wargnier-Dauchelle, V., Simon-Chane, C., and Histace, A. (2019). Retinal Blood Vessels Segmentation: Improving State-of-the-Art Deep Methods. In *Computer Analysis of Images and Patterns*, volume 1089, pages 5–16. -, Cham.

Yi, X., Walia, E., and Babyn, P. (2019). Generative Adversarial Network in Medical Imaging: A Review. *Medical Image Analysis*, 58:101552.