

Pair-GAN: A Three-Validated Generative Model from Single Pairs of Biomedical and Ground Truth Images

Clara Brémond-Martin^a, Huaqian Wu^b, Cédric Clouchoux^c and Kévin François-Bouaou^d

Witsee, 33 Av. des Champs-Élysées, 75008 Paris, France

Keywords: GAN, Single Input, Auto-Encoder, Biomedical, Pair, Segmentation.

Abstract: Generating synthetic pairs of raw and ground truth (GT) image is a strategy to reduce the amount of acquisition and annotation by biomedical experts. Pair image generation strategies, from single-input paired images (SIP), focus on patch-pyramid (PP) or on dual branch generator but, resulting synthetic images are not natural. With few-input images, for raw synthesis, adversarial auto-encoders synthesises more natural images. Here we propose Pair-GAN, a combination of PP containing auto-encoder generators at each level, for the biomedical image synthesis based upon a SIP. PP allows to synthesise using SIP while the AAE generator renders most natural the image content. We use for this work two biomedical datasets containing raw and GT images. Our architecture is evaluated with seven state of the art method updated for SIP: qualitative, similitude and segmentation metrics, Kullback Leibler divergences from synthetic and original feature image representations, computational costs and statistical analyses. Pair-GAN generates most qualitative and natural outputs, similar to original pairs with complex shape not produced by other methods, however with increased memory needs. Future works may use this generative procedure for multimodal biomedical dataset synthesis to help their automatic processing such as classification or segmentation with deep learning tools.

1 INTRODUCTION

Deep learning tools (DL) show powerful capabilities for image segmentation or classification in various computer-vision domains. The success of these tools, greedy on training data and resources, relies on the availability of large number of labeled images. In the case of biomedical images, imaging and annotations are difficult to obtain. Indeed the time required to replicates the experiments, or the resources constraints for ethical or sustainable reasons, hindrance the DL implementations.

Data augmentation is the method of choice to increase dataset images Xun et al. (2022); Iqbal et al. (2022). Classical transformations usually proposed, such as flip-flop or cropping, do not bring new content. Over the last decade, generative adversarial networks (GAN) have given a gather speed to the automation of image analysis supported by DL in the biomedical field Goodfellow et al. (2014). This

method synthesises a new specimen exhibiting the representative characteristics of the original images, however without duplicating one of them. The architecture sustaining this process is based upon two networks constituted by convolutional layers: a generator and a discriminator. The generator creates a new realistic sample aiming at misleading the second network, using an image or a noise as an input. The discriminator aims at determining if the image given in input is original or synthetic. Both networks are improving each other using a loss function. However the original GAN requires at least a minimal number of images in input to proceed and give significant results Lindner et al. (2019).

To overcome this limitation, many GAN dropping-down architectures have emerged to synthesise images with single input (SI) image. These architectures lie most of the time on pyramidal architectures or dual generators. In the first case, the generators of pyramidal single-input architecture are multiscale-patch, which consists in convolutional blocks taking in input a noise and an image at a specific level of resolution. The output of this resolution level is then given to the following block, also taking a noise and an image as inputs Shaham et al.

^a <https://orcid.org/0000-0001-5472-9866>

^b <https://orcid.org/0000-0003-1061-675X>

^c <https://orcid.org/0000-0003-3343-6524>

^d <https://orcid.org/0000-0002-9754-3065>

(2019). The second strategy consists in implementing a two-branch generator, one focusing on local and one on background features, with the same original GAN architecture Sushko et al. (2021a).

To the best of our knowledge, most of SI generative architecture are based upon GAN architecture, inheriting their limitations such as mode collapse or non-natural generations. Other solutions have appeared and could be integrated such as deep convolutions, semantic information or, a conditional input Iqbal et al. (2022). Only two studies in SI propose to replace the GAN generative convolutional layers by variational autoencoders, which helps synthesising diversified raw images Gur et al. (2020); Yoon et al. (2022). In previous work, some drawbacks has been observed in using the original GAN architecture to generate diversified and natural raw images, but based only on a reduced input images dataset. This GAN architecture replacement by an auto-encodeur improves the image generation and has been validated with three strategies Brémond Martin et al. (2022). However, the generation of a pair of images and corresponding ground truth (GT) in SI with an adversarial auto-encoder generator structure has not been tested yet.

Recently, two architectures have been updated to handle the generation of a pair of raw and annotated images from a single input pair (SIP) of raw and GT images. The pyramidal improvement has been developed to handle the GT by increasing the number of input channels Shaham et al. (2019). Another model addressing this issue lies in a dual generator from a noise a branch generates the raw image while the second synthesise the mask. An attention mechanism allows then to evaluate the realism of images Sushko et al. (2023).

The characterisation of the natural aspect of synthesised biological images has not yet been addressed. In the literature, there is no consensus on a metric or a metric combination allowing to ensure the naturality of an image, which is a real issue in the biomedical field Borji (2019). Recently, Brémond-Martin et al. (2023) propose to compare metric and psychovisual evaluation to choose appropriate metrics for an applicative case.

In this paper, we attempt to answer these questions by proposing a new Pair-GAN architecture using pyramidal auto-encoders generation for biomedical image and GT generations from a SIP. The pyramidal structure helps to generate images from SI and the auto-encoder more natural samples. Some improvement has been dedicated to the generative part of a SI image. However none of them identified if the concomitant generation of a raw and ground truth

images give similar results, and none of them to our knowledge use a pyramidal auto-encoders generation. We propose to compare the resulting synthetic pair from Pair- with synthetic pair from state of the art architectures dedicated to SI synthesis and adapted for pair generation. To avoid the lack of consensus in metrics, we use previously developed and validated metrics and statistical strategy Brémond Martin et al. (2022); Brémond-Martin et al. (2023) and a computational validation.

2 RELATED WORKS

2.1 Single Input Generative Methods

Pyramid Frameworks. To generate images from a SI image, the first strategy is to use pyramidal architectures. The model learns internal statistics from image patches across different image scales with a Wasserstein loss. Each stage is constituted by convolutional layers. SinGAN is the first model to synthesise images with a SI image strategy Shaham et al. (2019). Improvements of this architecture may rely on an attention module (SetGAN) or Gaussian smoothing Mahendren et al. (2023). Another recent model, ConSinGAN Hinz et al. (2021), is based upon a cascaded framework of SinGAN. In this model, a hierarchical module is added with a mixed reconstruction loss, providing various intensities of image generation, contrary to SinGAN. The Multi-scale-GAN lies on cascaded frameworks such as SinGAN. Similarly, it proposes an input based on noise and input images at various scales, the main difference being the use of PAC-Bayes boundary theory to tighter generalisation error bounds and synthesise more realistic images with a super-resolution Tang et al. (2022). SinIR uses the same SinGAN cascaded architecture with particular random pixels shuffling inside the generative part Yoo and Chen (2021). Likewise, Shuffling-SinGAN proposes a pixel shuffling at each scale of the pyramid generative framework, but adds a channel attention module and a spatial attention module. Shuffling pixels allows strengthening the role of the generator, misleading the discriminator Zheng et al. (2021a). While previous methods focus on textural information, SaSinGAN proposes to estimate larger image feature information by positioning attention modules at different locations according to the pyramid scale Chen et al. (2021b). CCASinGAN proposes cascades where the input characteristics are sketched into weighted feature maps, increasing the robustness of the attention module. Contrary to SaSinGAN where the attention modules are updating their posi-

tion at each scale, this architecture gives two attention modules at each scale around convolution layers Wang et al. (2022). Another multi-scale architecture consists of using Markov chain Monte Carlo as energy based model with no need of a discriminator network, as demonstrated in PatchGenCN Zheng et al. (2021b).

Discriminator Optimisations. One-shot-GAN, also known as SIV-GAN, focuses on a double discriminator module with a content and a layout branch to take into account spatial information and improve one shot image generation. This allows to not memorise previous training samples and to generate a different content from the origin Sushko et al. (2021b). The InGAN is constituted of a multi-scale discriminator composed of fully-convolutional patch discriminators. This formation allows capturing at each scale specific size patch statistics, coarse structures and details thanks to the multiscale discriminator Shocher et al. (2018). MorphGAN uses a double discriminator containing global and patch discriminators with a generator constituted with style encoder Ruiz et al. (2020).

Generator Optimisations. The Generalised One-Shot-GAN focuses on the generator optimisations and is composed of a main generative path and auxiliary branch inherited from the previous generator. Auxiliary branch aims to give assets of particular elements such as an object worn by a person, while the main branch focuses on the style, i.e. generating a person Zhang et al. (2022a). MoGAN architecture choose the same strategy by creating region of interest and background generator branches while the discriminator consists in Markovian chains. The ROI branch uses an affine transformation after convolutional layers to precise the synthesis Chen et al. (2021a). To maintain diversity and avoid collapse generation, HP-VAE-GAN uses hierarchical patch with VAE Gur et al. (2020). Similarly, the Our-GAN employs vertical coordinate convolutions to produce more natural contents Yoon et al. (2022). The RcGAN takes into account a random patch in generator input and consists in an cGAN architecture where the output of the antepenultimate layer takes the conditional vector Arantes et al. (2020).

Both Optimisations. For SGAN the generator and discriminator are following a DCGAN architecture in order to obtain better spatial information Jetchev et al. (2016). GenDa optimises the generation by adding a classifier in the discriminative part and truncate the

latent distribution of the generator with a strength factor Yang et al. (2021). PetsGAN avoids all the multi-stage construction and allows a one-step training with internal and external prior knowledge for the generative path and a regularised latent variable model. The internal priors restore high-fidelity textural information in images, and the external priors give the high-diversity and layout Zhang et al. (2022b)

GAN Combinations. A particular One-Shot image generation lies in the combination of many GAN frameworks such as Ex-Sin-GAN. This framework proposes a three-module assembly of GAN, each one focusing on either structural information, semantic, or textural information. While textural information is retrieved using SinGAN framework, the structural module is based upon a fully connected discriminator with a Wasserstein loss. Semantic information retrieval is based on a GAN inversion with a patch discriminator and a perceptual loss, where GAN inversion aims at finding the latent space code in a pre-trained way in order to best reconstruct images, giving the semantic content with the two other updates Zhang et al. (2021). A particular implementation lies in the AdvGAN framework which adds a perturbation after the generator, to render more diversified generation called Image to Image Translation Zhang (2019).

2.2 Pair Single-Input Generative Networks

To our knowledge, the only works generating a pair of raw and segmented images are from Shaham et al. (2019) and Sushko et al. (2023). The first one consists in the SinGAN architecture updated for four-channel images, with the last channel corresponding to the segmentation mask. The second work is based upon the One-shot architecture, however, a mask branch is added inside the generator and an attention module added in the discriminator.

These two architectures achieve to obtain similar images to the original pairs. The main drawback of SinGAN lies in the generator containing simple convolutional blocks, not allowing the synthesis of natural content. The One-Shot architecture proposes to add an attention module to overcome this issue. The strategy we choose to implement lies in the simple update of generator convolutional blocks without module additions, as an intent to render more natural the generative process.

3 METHODS

3.1 Datasets

We select two different biomedical datasets. The labeled and segmented gastrointestinal polyp images are from the HyperKvasir open source dataset, acquired with a standard endoscopy equipment from Olympus (Olympus Europe, Germany) and Pentax (Pentax Medical Europe, Germany). The second dataset consists of a labeled brain mice histopathological dataset, acquired in bright-field from the open microscopy project (idr0018-neff-histopathology/experimentA).

3.2 Resources

The scripts we create for this study are in python 3.11.3 with an Anaconda framework containing PyTorch 2.0.1 and cuda 117. We execute the codes on an Intel Core i7-10750H CPU, with 2.59 GHz and a Nvidia GeForce GTX 1650 TI GPU device.

3.3 Pair-GAN Framework

Our objective is to propose a hybrid architecture between multi-scale pyramidal patch and auto-encoders (AE) to generate images from a SIP of images. The natural diversity of the generation is given by the AE and, the statistics of a complex single image structures at various scales are captured by the multi-scale architecture.

3.3.1 Multi-Scale Architecture

Our proposed multi-scale architecture illustrated in Figure~1 consists of a patch-GANs pyramid, where each level is responsible for capturing the patch distribution at a different scale of input Shaham et al. (2019).

The pyramid starts at the lowest resolution and end with the finest resolution with a Gaussian noise image injected at each input. The deciphering is made with the corresponding original pair of image resolution. The full script is adapted such as the four channel images in SinGAN-Seg; the input is composed of the RGB raw image concatenated to the segmented image. The pyramid starts by giving a Gaussian image to a generator (G_n) which aims to map a noise (z_n) to a pair of raw and segmented images (pn):

$$pn = G_n(z_n) \quad (1)$$

The discriminator (D_n) attempts then to decipher if the pair of images given are original or synthetic. The

Table 1: Generator based upon an auto-encoder architecture. The input shape in pyramid for an image of 100×100 pixels are from the lowest to the highest resolution level: 25×25 , 33×33 , 44×44 , 58×58 , 76×76 and 100×100 . f: filter, k: kernel, LR: LeakyRelly, m: momentum, s: stride.

name	shape	parameters
conv	input level shape	(f=4, k=3, s=1, LR=0.2)
conv	max level shape	(f=32, k=3, s=1, LR=0.2)
Batchnorm		(m=0.1, affine)
deconv	latent space shape	(f=32, k=3, s=1, LR=0.2)
deconv	max level shape	(f=4, k=3, s=1, LR=0.2)
Batchnorm		(m=0.1, affine)
conv	output level shape	(f=1, k=3, s=1, Tanh)

two networks update each other by an adversarial loss. The single difference between the G_n and the next levels [$G_{n-1} : G_0$] is that [$G_{n-1} : G_0$] receive in addition to the Gaussian image the raw and segmented pair of images at a finest resolution, thus:

$$pn - 1 = G_{n-1}(zn - 1(pn_{upsampled})) \quad (2)$$

Generator. In the literature, the Generator of PathGAN or SinGAN and its derivatives is based upon a classical GAN approach containing only convolutional blocks Shaham et al. (2019). Here we use an AE generator architecture as detailed in Table 1 Brémond Martin et al. (2022); Brémond-Martin et al. (2023). The aim of this architecture is to reconstruct a pair of raw and segmented image from a latent space s composed of an encoder e , and a decoder d . The encoder transforms the input by applying one or more nonlinear parameters into a new representation of a lower dimension. The decoder also uses nonlinear transformations to reconstruct the original image with the lower dimensional representation.

Training. The chosen discriminator is the same Markovian discriminator than the Patch-GAN, SinGAN and SinGAN-Seg implementation Shaham et al. (2019). The selected loss reconstruction is from SinGAN implementation. Instead of using a gradient penalty to optimise the Wasserstein loss, we use Perceptual-Wasserstein loss from our previous study to reproduce similar condition and a better contrast and natural images Brémond Martin et al. (2022); Brémond-Martin et al. (2023). This loss is applied on the whole images and not on patch to help the network to better learn the boundaries. This architecture is a Vanilla kind, at the end of the pyramid training, the 40 pair of images generated from a single original pair are retrieved for comparisons.

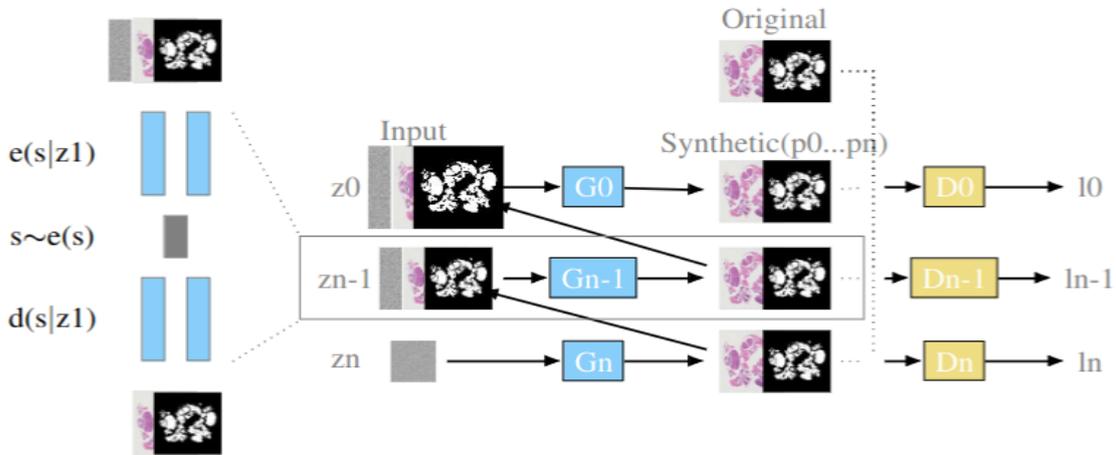


Figure 1: Pair-GAN for single input pair generation of raw and segmented biomedical images. This architecture contains the classical cascaded network observed in the literature for the single input generation, hybridised with an auto-encoder replacement instead of simple convolutions at each scale to give a natural content. z_1 stands for input, s for the latent space, $e(s|z_1)$ for an encoding distribution and $d(s|z_1)$ for the decoding distribution.

3.4 Evaluations

3.4.1 State of the Art Comparisons

To verify our results fairly, we choose to compare the resulting raw and segmented pairs from our architecture with pairs resulting from the run of others architectures on the same datasets: SinGAN Shaham et al. (2019), ConSinGAN Hinz et al. (2021), hp-VAE-GAN Gur et al. (2020), SIV-GAN Sushko et al. (2023), InGAN Shocher et al. (2018), PetsGAN Zhang et al. (2022b) and Ex-Sin-GAN Zhang et al. (2021). The scripts have been adapted to take as input a four-channel image corresponding to an RGB image and its GT except for SinGAN and Siv-GAN which propose input pairs solution. For each architecture, forty pairs of images are generated from each pair of input images. For a simple visualisation, we show a simple sample of input image, and a single sample pair of each forty pairs of images resulting from each tested architecture. We colored on segmented samples the added (pink) or eliminated (green) region of interest (ROI) during the generation to verify the diversity from the original GT. To observe the forty generated masks shape and variations at the same time, a heatmap is produced where the most generated pixels are in pink and the background in black.

In order to verify the interest of using in the generative part the adversarial auto-encoder in single input, we compare the AAE generation Brémond Martin et al. (2022) with a GAN Goodfellow et al. (2014), a DCGAN Wu et al. (2020), and an INFO-GAN Chen et al. (2016). To verify the number of image contributions during the generation, we test various few input synthesis: 20, 15, 10, 5 images in input and then

the single input configuration. To estimate the pair generation interest in the same conditions, we synthesise raw images and then ground-truth separately in a vanilla way and stop the generation at 2000 epochs see Tables 8, 7, 9 and 3.

3.4.2 Dimensional Reduction

To verify if the generated raw and segmented images are in the same feature space than the original image, we apply a t-SNE dimensional reduction analysis. In order to compare the groups constituted by images synthesised from each architecture, the mean Kullback-Leibler distance is evaluated. First it has been calculated between the representations of original and synthetic pair of images, then between the original sample tested and all the other original samples contained inside a dataset.

3.4.3 Metrics

Due to the lack of consensus for natural synthetic image characterisation, various metrics are evaluated to give an asset of the similitude with the original pair of images, and give their quality Brémond Martin et al. (2022); Borji (2019). All raw and segmented pairs are compared in terms of: Blur, structural similitude index (SSIM), mutual information (MI), peak signal-to-noise ratio (PSNR), mean square error (MSE), universal quality metric (UQM), single image Fréchet inception distance (SIFID) and learned perceptual image patch similarity (LPIPS).

For segmented images, the interest is to consider if the generative networks are generating more, less, similar or various segmentation shapes than the original dataset. For this reason, we calculate Jaccard,

area of segmentation, differences between the original GT and a synthetic segmentation, accuracy, sensitivity, specificity, dice and f1 scores. The generation giving the most diversified content different from the GT is expected. All the mean, median, minimum, maximum and standard deviation are given for each metric and each group of architecture.

3.4.4 Statistics

The result of metrics (8 variables: Blur, SSIM, MI, PSNR, MSE, UQM, SIFID, LPIPS and then 8 variables: Jaccard, Area, Ori-syntheticarea, Accuracy, sensitivity, specificity, dice, f1) by group of architectures (8 factors: SinGAN, ConSinGAN, hp-VAE-GAN, SIV-GAN, InGAN, PetsGAN, Ex-Sin-GAN and Pair-GAN) are then compared statistically, and also against the original dataset (9th factor: Original). For this reason, we choose a Kruskal-Wallis method followed by a Conover post-hoc in order to compare all results from all the architectures against the original pairs of images. Alpha risk is considered at 5%.

3.4.5 Computational

To verify the benefit of using such methods, we calculate the execution time (in seconds), and memory usage (in Tb).

4 RESULTS

The objective here is to compare in various ways and on various datasets if the hybridisation of an AE with a multi-scale generative network helps to generate more natural and diversified images and their GT with a SIP of biomedical images. In this part, we compare qualitatively, with metrics, statistics, and computationally, the results of our generative architecture with state-of-the-art results and the original pairs of images.

4.1 Qualitative Comparisons

We compare visually datasets containing GT labeled images with the synthetic raw and segmented images from all the architecture. In Table 2, a sample experiment from a SIP of images is presented. For the polyp dataset, generated raw images have the same aspect, color and brightness as the original image regardless of the method. The only aspect diverging in synthetic raw results is the red coloration in the middle of the polyp which disappears in SinGAN, ConSinGAN and InGAN results while, some red pigmentations are observed in our Pair-GAN architecture.

Concerning the segmented mask, Siv-GAN, PetsGAN or Ex-Sin-GAN does not create a particular shape variation compared to the original dataset. As shown in the example, some architectures are creating structures with less region of interest (ROI) such as SinGAN, ConSinGAN, INGAN or our Pair-GAN. Only hp-VAE-GAN creates an extension of structure, adding new ROI.

The heatmap representation of all the segmentation mask generated by each architecture depicts almost a similar ROI synthesis for each GAN.

For histological dataset, with complex structures, if the mice brain color and space representation is reproduced by the GANs networks, only the Pair-GAN architecture seems to generate all the structures. Others solutions are mainly generating similar structures at peripheral zones of the microscopic acquisition.

Whatever the architecture used, almost the same proportion of additional or removal ROI are presented in synthetic segmentation mask.

The heatmap renders external brain structure zones more contrasted than the internal brain zones, except for Pair-GAN images. Our solution visually generates the most similar and robust generation, whatever the brain structure considered, with fewer variations.

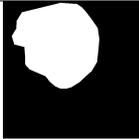
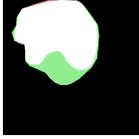
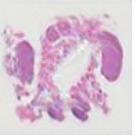
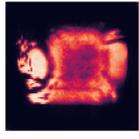
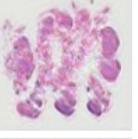
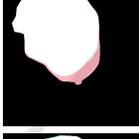
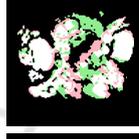
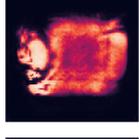
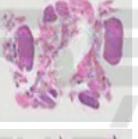
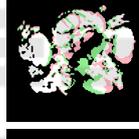
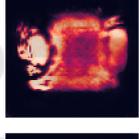
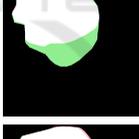
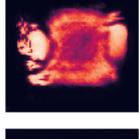
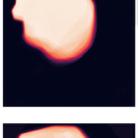
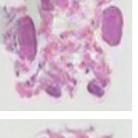
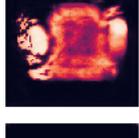
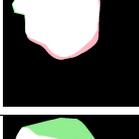
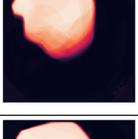
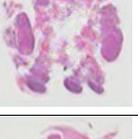
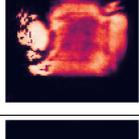
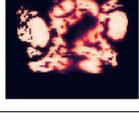
We propose in the following subsection to verify our observation by statistically comparing metrics and computational calculations on the overall datasets.

4.2 Statistical Space Comparison

In this part, we study the representation of original and synthetic pairs of images in the same optimised statistical space. Dimensional reduction is applied to the extracted features on images during the generative process. We then calculate Kullback Leibler divergence (KL) on the ten t-SNE representations to verify the stability of the representation shown in Figure 2.

To verify which architecture generates the most diversified representation compared with the original single pair of images, the mean KL between the original representations and the generated raw or segmented synthetic image representations is calculated. For the polyp dataset, Pair-GAN generates raw images that are the most divergent from the original raw images, followed by INGAN and CONSIN. The mean KL between GT and segmented representations is greater than the mean KL between raw and synthetic image representations. For this kind of images, Pair-GAN and CONSIN reach the highest scores and give the most divergence from the original GT image. According to the histological dataset, the KL divergence is reduced when original and Pair-GAN syn-

Table 2: Biomedical and segmented synthetic data paired generated with a SIP image with various GANs. In green are represented synthetic missing regions and in pink added regions compared to the original GT image. The heatmap are representing all the generated sample segmentation with the most occurring segmentation in pink and in black the pixels which are never corresponding to a segmented region.

Dataset	Original	GT		Original	GT	
Model	Synthetic	Segmented	Heatmap	Synthetic	Segmented	Heatmap
						
SinGAN						
ConSinGAN						
hp-VAE-GAN						
One-Shot-GAN						
INGAN						
Pets-GAN						
Ex-Sin-GAN						
Pair-GAN(ours)						

thetic images are considered. For the raw synthesis, the Ex-Sin-GAN generation gives the second-lowest divergence and, for the segmentation, the ONE-SHOT generation.

To verify if these divergences are too far from natural images, we then compare original and synthetic pairs in terms of metrics.

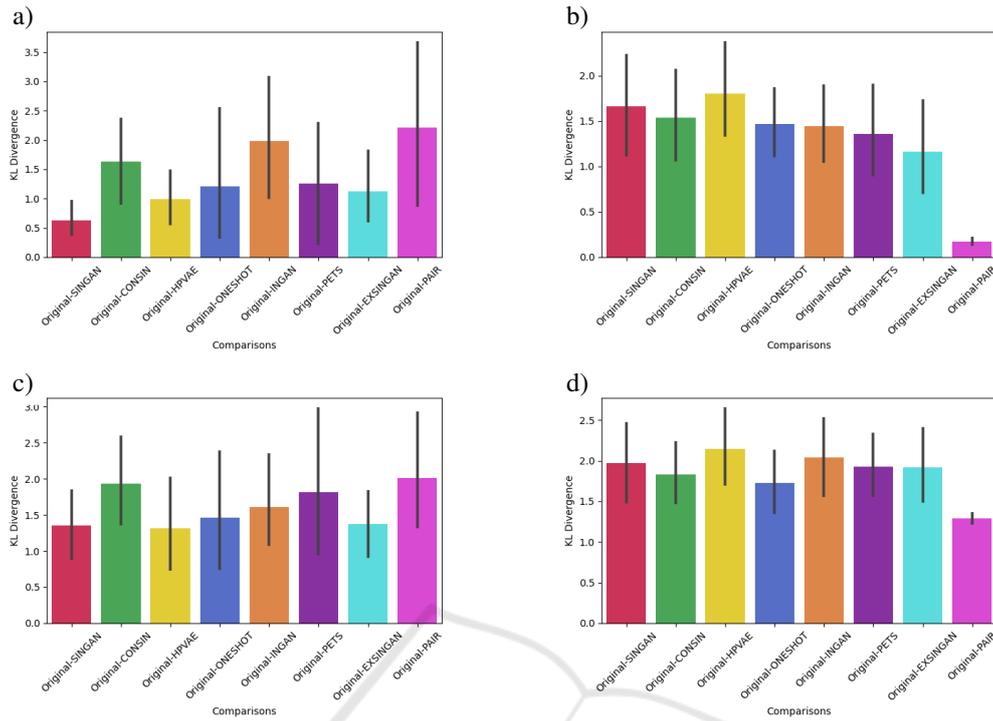


Figure 2: KL divergences on t-SNE on synthetic raw and segmented images in the same space as an original polyp image. a) Raw polyp dataset b) Raw histological dataset c) Segmented polyp dataset and d) Segmented histological dataset.

4.3 Metric Comparisons

4.3.1 Qualitative and Similitude Metrics

The summary of metric calculations is presented in Table 3.

For the Polyp dataset, raw synthetic images are different from the original raw images, whatever the architecture and metric considered. No differences can be observed between synthetic raw images from Pair-GAN and other generations.

For the histological dataset, raw images are not different from original images, except if the generation is executed with Pair-GAN. This generation differs from other generative architecture, whatever the metric considered ($p < 0.001$). Concerning the segmented images, only differences between PAIR-GAN and HP-VAE-GAN can be observed see Appendix Table 6.

Thus, for simple rounded objects in biological images, the synthesis is qualitative and similar (low PSNR high SSIM, MI, UQI) whatever the architecture. However, for complex shapes, the synthesis produces a different content from the original image (high SSIM, MI, UQI scores) and loses some quality (increase of PSNR scores) except with Pair-GAN.

For the raw polyp dataset the blurriest images are synthesised with Consin (with 1253.98), the noisiest

with Ingan (with 19.13) and the best similitude scores depends on the architecture.

For the raw histological dataset, the blurriest images are synthesised with HP-VAE-GAN and Consin (with 27.50 and 27.08), the noisiest and the most similar with Pair-GAN (Psnr of 23.90, lowest MSE of 287.96 for quality and, highest Ssim of 0.73, lowest SIFID of 236.94 and LPIPS of 0.29 for the similitude).

4.3.2 Segmentation Metrics

Here we propose to use segmentation metrics to verify the differences of synthesis segmentation compared to the GT and to better characterize the generation. Indeed, previously we observed the generation can add or eliminate some regions of interest depending on the architectures, these metrics could help to quantify these differences and are summarized in Appendix Table 5.

The Pair-GAN Jaccard scores for segmented polyp images tend toward the Original dataset variation. All the other scores do not seem to be relevant to compare, as the comparison between original input sample and others originals are far. Indeed, the f1 score between a GT used in input and all the others GT in the original dataset is 0.28, dice at 0.37, sensitivity at 0.24, accuracy at 0.66 and a difference area

Table 3: Metrics calculated on one-shot segmented synthetic images from various architectures. The stars correspond to significant differences between a group of generation with the corresponding input with *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. Best scores tending to perfect scores for a quality or similitude metric are bolded.

Metrics part A									
	Blur		Mi		SSIM		Psnr		
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	
SI Vs All Original	1361.07	546.48	1743.76	56.82	0.24	0.03	11.87	3.00	
Polyp	SinGAN	1011.16 ***	106.80	1.06 ***	0.01	0.29 ***	0.02	18.62 ***	0.60
	CON SIN	1253.98 **	89.21	1.13 **	0.03	0.33 **	0.02	18.70 **	0.56
	HP-VAE-GAN	576.02 ***	46.50	1.13 ***	0.07	0.35 ***	0.02	18.76 ***	0.99
	ONE-SHOT-GAN	687.90 ***	45.20	1.12 ***	0.04	0.34 ***	0.02	19.00 ***	0.68
	INGAN	1058.21 ***	65.43	1.15 ***	0.04	0.35 ***	0.02	19.13 ***	0.77
	Ex-Sin-GAN	938.56 ***	80.43	1.09 ***	0.04	0.32 ***	0.02	18.47 ***	0.67
	PETSGAN	1154.50 **	73.30	1.16 **	0.04	0.35 **	0.02	19.00 **	0.60
PAIRGAN (Ours)	1055.14 ***	77.67	1.11 ***	0.04	0.29 ***	0.02	18.68 ***	0.59	
SI Vs All Original	46.47	6.64	-1824.43	303.29	0.41	0.14	19.71	18.43	
Histological	SinGAN	25.52	2.09	-1810.91	20.99	0.52	0.03	19.06	0.73
	CON SIN	27.08	1.40	-1802.24	23.29	0.53	0.04	19.31	0.82
	HP-VAE-GAN	27.50	1.97	-1820.40	19.12	0.51	0.03	18.78	0.62
	ONE-SHOT-GAN	26.91	1.59	-1800.55	25.42	0.54	0.04	19.31	0.80
	INGAN	26.50	1.58	-1807.53	21.94	0.53	0.03	19.28	0.76
	Ex-Sin-GAN	25.10	1.75	-1798.80	23.57	0.55	0.04	19.35	0.83
	PETSGAN	23.35	1.59	-1795.49	22.48	0.56	0.037	19.79	0.82
PAIRGAN (Ours)	17.33 ***	2.89	-1677.16 ***	45.82	0.73 ***	0.07	23.90 ***	1.80	
Metrics part B									
	MSE		UQM		SIFID		LPIPS		
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	
SI Vs All Original	4540.44	1469.52	0.72	0.08	890.92	117.81	0.57	0.03	
Polyp	SinGAN	902.60 ***	128.73	0.91 ***	0.01	757.64 ***	90	0.46 ***	0.01
	CON SIN	885.50 **	113.84	0.91 ***	0.01	623.48 **	114.49	0.43 **	0.01
	HP-VAE-GAN	889.90 ***	221.96	0.92 ***	0.01	757.13 ***	100.60	0.42 ***	0.01
	ONE-SHOT-GAN	828.93 ***	140.34	0.92 ***	0.01	778.04 ***	106.62	0.44 ***	0.01
	INGAN	808.47 ***	153.73	0.92 ***	0.01	637.66 ***	99.32	0.40 ***	0.01
	Ex-Sin-GAN	935.92 ***	154.23	0.91 ***	0.01	759.73 ***	139.39	0.44 ***	0.01
	PETSGAN	826.10 **	121.79	0.92 ***	0.01	511.43 **	90.14	0.40 **	0.01
PAIRGAN (Ours)	889.18 ***	123.61	0.91 ***	0.01	598.45 ***	92.15	0.44 ***	0.01	
SI Vs All Original	1756.86	436.67	0.96	0.01	377.96	106.28	0.54	0.13	
Histological	SinGAN	817.77	136.27	0.98	0.01	465.04	95.37	0.45	0.02
	CON SIN	775.76	143.97	0.98	0.01	460.51	77.29	0.44	0.03
	HP-VAE-GAN	869.64	123.48	0.98	0.01	416.30	98.95	0.46	0.02
	ONE-SHOT-GAN	775.42	135.63	0.98	0.01	382.49	77.83	0.44	0.02
	INGAN	779.18	133.07	0.98	0.01	466.02	112.45	0.44	0.03
	Ex-Sin-GAN	769.15	146.74	0.98	0.01	480.09	109.76	0.44	0.03
	PETSGAN	694.36	125.28	0.98	0.01	414.39	99.71	0.44	0.02
PAIRGAN (Ours)	287.96 ***	115.51	0.99 ***	0.01	236.94 ***	100.81	0.29 ***	0.04	

around 11000 while the difference between the GT with the synthetic reaches are different, see Appendix Table 5. Thus, for the area, difference area, accuracy, sensitivity, specificity, dice and F1-score, the Pair-GAN result is different from the Original ($p < 0.001$). There are no statistical differences between the Pair-GAN architecture segmentation synthesis and other architecture results, except for the Jaccard and area scores (and particularly with the HP-VAE-GAN architecture with $p < 0.001$).

Histological synthetic segmentation is different from the original GT for all the architectures accord-

ing to all the scores except for the accuracy with Pair-GAN architecture (0.78 versus 0.77, $p > 0.05$). The Pair-GAN dice and f1 scores are significantly weaker than other architectures ($p < 0.01$ while for others groups $p < 0.001$). The segmentation synthesis is different for all the scores with Pair-GAN and others architectures, except for the areas (the only difference concerns Pair-GAN with CON SIN $p < 0.001$), the total area scores (the only difference concerns Pair-GAN with ONE-SHOT $p < 0.01$), and the Jaccard (there are no differences between Pair-GAN and PETS-GAN $p > 0.05$).

Table 4: Computational comparisons.

Architecture	Time(s)	Memory (Tb)
SinGAN	1944.62	2986.95
ConSinGAN	2047.64	2871.04
hp-VAE-GAN	2020.55	2452.71
One-Shot-GAN	2245.21	582.03
INGAN	1999.47	2482.88
Pets-GAN	3747.69	441.43
Ex-Sin-GAN	8377.86	2785.37
Pair-GAN	2020.25	3034.33

4.4 Computational Comparisons

Pair-GAN run in the same time laps than SinGAN architecture see Table 4(2020s vs 1944s). Concerning memory consumption, Pair-GAN is the most expensive (3034Tb) and Pets-GAN needs the less memory. The Ex-Sin-GAN architecture needs almost four times the SinGAN time to be executed.

5 DISCUSSION

We present Pair-GAN, a hybrid framework containing a multi-scale architecture with auto-encoders to help generation of pairs of raw and segmented biomedical images from a SIP. This generation gives natural images and an accurate segmentation, which are considered as qualitative and similar to the original SI from which they have been generated. They are also in the same statistical space as the original dataset. Contrary to other frameworks, Pair-GAN generates higher diversity from a SI as shown in the dimensional reduction and the metric calculation for the polyp dataset. For the histological dataset, with more complex shape, it renders the most similar and natural representations. However, Pair-GAN requires a huge need of memory to be executed in approximately the same time as other networks.

To our knowledge, Pair-GAN is the third implementation dedicated to a generation pair from a SIP (SinGAN and ConSin-GAN) Shaham et al. (2019); Sushko et al. (2023). This strategy generates images as natural as the other two frameworks, nevertheless with more similar structures than SinGAN and ConSin-GAN. These interesting results may be due to the auto-encoding structure we added in the pyramid layer combined with the perceptual loss Brémond Martin et al. (2022) and Appendix Tables 8, 7, 9, 3. To verify these observations, a psychovisual study of raw synthetic images may help to decipher if these images are also considered as natural by biological experts, and particular architecture misleads more experts than others such as in Brémond-Martin

et al. (2023). Another interesting approach to verify the segmentation is trying to classify these images (by their physio-pathological content, for instance). Here we observe the results on two datasets with binary segmentation, an interesting project would be to test the generation of multimodal segmentation in order to produce a data augmentation dedicated to multi-label classifications Pandeva and Schubert (2019); Hong et al. (2020).

Pair-GAN generates more diversified contents than other architectures, as seen in the statistical space comparison for simple shape. This diversity could be a representation of the natural diversity present in the original dataset composed of multiple labeled images. This diversity of generation is a strong feature researched in GAN field in order to improve deep learning training for image segmentation for instance Xun et al. (2022). For a qualitative segmentation, increasing the number of natural samples with various features allows to render accurate the results. However, more natural and diversified augmented samples in a segmentation training process, without augmenting the total number of training images allows to precise the segmentation Brémond-Martin et al. (2023). An interesting project may be to verify which data augmentation architecture with single or few-input renders the most accurate segmentations.

Our architecture gives the most similar results (SSIM,UQM,SIFID for instance) for complex biological contents. For the histological dataset, only Pair-GAN seems to produce a realistic image despite the complexity of histological multiple structures. Indeed, minima and maxima similitude scores for segmented images from Pair-GAN are near the minima and maxima GT scores. It is not the case for other architectures, and these metrics results may be linked with heatmap observations. For the polyp dataset, containing a single object, all architectures except Pair-GAN generate images similar to the input original images. Pair-GAN generates shapes not already produced (none of the similitude scores are reaching their maximum) as shown in Table 3. Thus, a set of metrics seems to help verify the naturality of images produced with our architecture, while there is currently no consensus on the use of a particular metric to validate the GAN synthesis Borji (2019). Another interesting project should be to verify if these metrics are really useful in psychovisual evaluation task Brémond-Martin et al. (2023). Particularly, if they help to identify the naturality of rounded or simple (such as polyp) or multiple or complex segmented objects (such as brain structures) in images.

Pair-GAN does not require more time to be executed, but needs a consequent amount of memory.

The addition of an encoding and decoding part with latent space increases memory needs. In comparison, the original SinGAN structure, composed of only convolution layers, needs less memory Shaham et al. (2019). To reduce the memory requirement, a future work is to lighten the encoding architecture and optimise the latent space. Instead of using the maximal resolution in the second encoding convolutional layer, an idea could be to update the shape according to the level of resolution, which may accurate the results.

In future directions, other noise inputs may be evaluated for the generative part. Indeed, in previous research, we show an influence of the noise injection according to a kind of acquisition Brémond Martin et al. (2022). Thus, it could be interesting to update the noise injection according to the microscopic acquisition considered and test if the result from a particular injection is still linked with the acquisition. For the polyp dataset, the update of noise input may reproduce particular saturation, the over/under exposure of polyp topology during the imaging Ali et al. (2020). Additionally, the generated image specularities may be evaluated to enhance the generated light reflection on smooth objects. The loss function may also be improved. As shown previously, it can improve the contrast between the background and the researched structure Brémond Martin et al. (2022).

6 CONCLUSION

In this article we present Pair-GAN:

- A generative architecture based upon patch-pyramidal auto-encoders;
- Taking in input a single pair of raw and GT biomedical images;
- Which synthesise natural images, similar and in the same statistical space as original pairs and compared with state-of-the-art methods.

Such approach may be interesting to increase minimal datasets to automate for instance the diagnosis grade of a disease from a single image with deep learning methods. An interesting perspective may be to verify the grade of each generation from a single input pair of images annotated with the grade.

REFERENCES

Ali, S., Zhou, F., Braden, B., Bailey, A., Yang, S., Cheng, G., Zhang, P., Li, X., Kayser, M., Soberanis-Mukul, R. D., et al. (2020). An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. *Scientific reports*, 10(1):2748.

Arantes, R. B., Vogiatzis, G., and Faria, D. R. (2020). Regan: learning a generative model for arbitrary size image generation. In *Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part I 15*, pages 80–94. Springer.

Borji, A. (2019). Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, 179:41–65.

Brémond-Martin, C., Simon-Chane, C., Clouchoux, C., and Histace, A. (2023). Brain organoid data synthesis and evaluation. *Frontiers in Neuroscience*, 17.

Brémond Martin, C., Simon Châne, C., Clouchoux, C., and Histace, A. (2022). Aaegan loss optimizations supporting data augmentation on cerebral organoid bright-field images. In *VISIGRAPP (4: VISAPP)*, pages 307–314.

Chen, J., Xu, Q., Kang, Q., and Zhou, M. (2021a). Mogan: Morphologic-structure-aware generative learning from a single image. *arXiv preprint arXiv:2103.02997*.

Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., and Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in neural information processing systems*, 29.

Chen, X., Zhao, H., Yang, D., Li, Y., Kang, Q., and Lu, H. (2021b). Sa-singan: self-attention for single-image generation adversarial networks. *Machine Vision and Applications*, 32:1–14.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Gur, S., Benaim, S., and Wolf, L. (2020). Hierarchical patch vae-gan: Generating diverse videos from a single sample. *Advances in Neural Information Processing Systems*, 33:16761–16772.

Hinz, T., Fisher, M., Wang, O., and Wermter, S. (2021). Improved techniques for training single-image gans. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1300–1309.

Hong, D., Yao, J., Meng, D., Xu, Z., and Chanussot, J. (2020). Multimodal gans: Toward cross-modal hyperspectral–multispectral image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):5103–5113.

Iqbal, A., Sharif, M., Yasmin, M., Raza, M., and Aftab, S. (2022). Generative adversarial networks and its applications in the biomedical image segmentation: a comprehensive survey. *International Journal of Multimedia Information Retrieval*, 11(3):333–368.

Jetchev, N., Bergmann, U., and Vollgraf, R. (2016). Texture synthesis with spatial generative adversarial networks. *arXiv preprint arXiv:1611.08207*.

Lindner, L., Narnhofer, D., Weber, M., Gsaxner, C., Kolodziej, M., and Egger, J. (2019). Using synthetic training data for deep learning-based gbm segmentation. In *2019 41st Annual International Conference of*

- the *IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6724–6729. IEEE.
- Mahendren, S., Edussooriya, C. U., and Rodrigo, R. (2023). Diverse single image generation with controllable global structure. *Neurocomputing*, 528:97–112.
- Pandeva, T. and Schubert, M. (2019). Mmgan: Generative adversarial networks for multi-modal distributions. *arXiv preprint arXiv:1911.06663*.
- Ruiz, N., Theobald, B.-J., Ranjan, A., Abdelaziz, A. H., and Apostoloff, N. (2020). Morphgan: One-shot face synthesis gan for detecting recognition bias. *arXiv preprint arXiv:2012.05225*.
- Shaham, T. R., Dekel, T., and Michaeli, T. (2019). Singan: Learning a generative model from a single natural image. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4570–4580.
- Shocher, A., Bagon, S., Isola, P., and Irani, M. (2018). Ingan: Capturing and remapping the” dna” of a natural image. *arXiv preprint arXiv:1812.00231*.
- Sushko, V., Gall, J., and Khoreva, A. (2021a). One-shot gan: Learning to generate samples from single images and videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2596–2600.
- Sushko, V., Zhang, D., Gall, J., and Khoreva, A. (2021b). Generating novel scene compositions from single images and videos. *arXiv preprint arXiv:2103.13389*.
- Sushko, V., Zhang, D., Gall, J., and Khoreva, A. (2023). One-shot synthesis of images and segmentation masks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6285–6294.
- Tang, J., Tao, B., Gong, Z., and Yin, Z. (2022). Adaptive adversarial training method for improving multi-scale gan based on generalization bound theory. *arXiv preprint arXiv:2211.16791*.
- Wang, X., Jiang, W., Zhao, L., Liu, B., and Wang, Y. (2022). Ccsingan: Cascaded channel attention guided single-image gans. In *2022 16th IEEE International Conference on Signal Processing (ICSP)*, volume 1, pages 61–65. IEEE.
- Wu, Q., Chen, Y., and Meng, J. (2020). Dcgan-based data augmentation for tomato leaf disease identification. *IEEE Access*, 8:98716–98728.
- Xun, S., Li, D., Zhu, H., Chen, M., Wang, J., Li, J., Chen, M., Wu, B., Zhang, H., Chai, X., et al. (2022). Generative adversarial networks in medical image segmentation: A review. *Computers in biology and medicine*, 140:105063.
- Yang, C., Shen, Y., Zhang, Z., Xu, Y., Zhu, J., Wu, Z., and Zhou, B. (2021). One-shot generative domain adaptation. *arXiv preprint arXiv:2111.09876*.
- Yoo, J. and Chen, Q. (2021). Sinir: Efficient general image manipulation with single image reconstruction. In *International Conference on Machine Learning*, pages 12040–12050. PMLR.
- Yoon, D., Oh, J., Choi, H., Yi, M., and Kim, I. (2022). Ourgan: One-shot ultra-high-resolution generative adversarial networks. *arXiv preprint arXiv:2202.13799*.
- Zhang, W. (2019). Generating adversarial examples in one shot with image-to-image translation gan. *IEEE Access*, 7:151103–151119.
- Zhang, Z., Han, C., and Guo, T. (2021). Exsingan: Learning an explainable generative model from a single image. *arXiv preprint arXiv:2105.07350*.
- Zhang, Z., Liu, Y., Han, C., Guo, T., Yao, T., and Mei, T. (2022a). Generalized one-shot domain adaptation of generative adversarial networks. *Advances in Neural Information Processing Systems*, 35:13718–13730.
- Zhang, Z., Liu, Y., Han, C., Shi, H., Guo, T., and Zhou, B. (2022b). Petsgan: Rethinking priors for single image generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3408–3416.
- Zheng, M., Zhang, P., Gao, Y., and Zou, H. (2021a). Shuffling-singan: Improvement on generative model from a single image. In *Journal of Physics: Conference Series*, volume 2024, page 012011. IOP Publishing.
- Zheng, Z., Xie, J., and Li, P. (2021b). Patchwise generative convnet: Training energy-based models from a single natural image for internal learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2961–2970.

APPENDIX

We implemented the baseline models for single input image with open sources we adapted to generation from various channel input. These are the links to the open sources we employed:

- SinGAN: <https://github.com/tamarott/SinGAN>;
- ConSinGAN: <https://github.com/tohinz/ConSinGAN>;
- HP-VAE-GAN: <https://github.com/shirgur/hp-vaegan>;
- SIV-GAN: <https://github.com/boschresearch/one-shot-synthesis>;
- InGAN: <https://github.com/Caenorst/InGAN/tree/py3>;
- PETS-GAN: <https://github.com/zhangzc21/petsgan>;
- Ex-Sin-GAN: <https://github.com/zhangzc21/ExSinGAN>

Table 5: Appendix segmentation metrics on segmented synthetic images compared to the single original segmented image. The stars correspond to *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Dataset	Architecture	jaccard		area		ori-syntheticarea		accuracy	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
	SI Vs all Original	0.25	0.01	31219.698	25196.51	11004.698	25196.51	0.66	0.09
Polyp	SinGAN	0.20 ***	0.01	52925.52 ***	8278.16	32710.52 ***	8278.16	0.94 ***	0.03
	CONSIN	0.21 ***	0.01	53483.52 ***	5871.89	33268.52 ***	5871.89	0.94 ***	0.02
	HP-VAE-GAN	0.21 ***	0.004	50288.1 ***	10491.23	30073.1 ***	10491.23	0.93 ***	0.05
	ONE-SHOT-GAN	0.21 ***	0.003	53970.6 ***	4427.57	33755.6 ***	4427.57	0.95 ***	0.02
	INGAN	0.22 ***	0.01	51999.66 ***	5983.92	31784.66 ***	5983.92	0.94 ***	.03
	Ex-Sin-GAN	0.22 ***	0.01	51243.18 ***	5967.71	31028.18 ***	5967.71	0.94 ***	0.03
	PETSGAN	0.22 ***	0.007	54153.54 ***	3598.14	33938.54 ***	3598.14	0.95 ***	0.01
	PAIRGAN	0.23 ***	0.009	53073.9 ***	5356.36	32858.9 ***	5356.36	0.95 ***	0.02
	SI Vs all Original	0.95	0.02	374394.0	32243.30	254031.0	32243.30	0.78	0.06
Histological	SinGAN	0.16 ***	0.01	60167.0 ***	3133.26	-60196.0 ***	3133.25	0.74 ***	0.01
	CONSIN	0.17 ***	0.01	60410.62 ***	2968.53	-59952.38 ***	2968.53	0.75	0.01 ***
	HP-VAE-GAN	0.18 ***	0.01	61367.82 ***	3512.00	-58995.18 ***	3512.00	0.75 ***	0.01
	ONE-SHOT-GAN	0.18 ***	0.01	61623.28 ***	3669.62	-58739.72 ***	3669.62	0.75 ***	0.01
	INGAN	0.19 ***	0.01	61246.16 ***	3362.73	-59116.84 ***	3362.73	0.75 ***	0.01
	Ex-Sin-GAN	0.19 *	0.01	61790.96 ***	3268.68	-58572.04 ***	3268.68	0.75 ***	0.01
	PETSGAN	0.19 ***	0.01	62139.4 ***	2735.36	-58223.6 ***	2735.36	0.75 **	0.01
	PAIRGAN	0.18 ***	0.01	61135.9 ***	1568.16	-59227.1 ***	1568.16	0.77	0.01
	SI Vs all Original	0.24	0.22	0.87	0.13	0.37	0.22	0.28	0.21
Polyp	SinGAN	0.84 ***	0.11	0.98 ***	0.02	0.82 ***	0.02	0.89 ***	0.07
	CONSIN	0.85 ***	0.080	0.99 ***	0.01	0.83 ***	0.01	0.91 ***	0.04
	HP-VAE-GAN	0.81 ***	0.16	0.99 ***	0.01	0.81 ***	0.05	0.88 ***	0.12
	ONE-SHOT-GAN	0.87 ***	0.06	0.99 ***	0.006	0.83 ***	0.01	0.92 ***	0.04
	INGAN	0.84 ***	0.08	0.99 ***	0.01	0.83 ***	0.01	0.90 ***	0.05
	Ex-Sin-GAN	0.83 ***	0.09	0.99 ***	0.006	0.82 ***	0.01	0.90 ***	0.05
	PETSGAN	0.88 ***	0.05	0.99 ***	0.01	0.83 ***	0.01	0.92 ***	0.03
	PAIRGAN	0.86 ***	0.07	0.99 ***	0.009	0.83 ***	0.01	0.91 ***	0.04
	SI Vs all Original	0.78	0.05	0.78	0.08	0.73	0.04	0.77	0.06
Histological	SinGAN	0.47 ***	0.02	0.98 ***	0.01	0.48 ***	0.01	0.63 ***	0.02
	CONSIN	0.47 ***	0.02	0.98 ***	0.01	0.48 ***	0.01	0.63 ***	0.02
	HP-VAE-GAN	0.48 ***	0.03	0.97 ***	0.01	0.48 ***	0.01	0.63 ***	0.02
	ONE-SHOT-GAN	0.48 ***	0.03	0.98 ***	0.01	0.48 ***	0.01	0.64 ***	0.03
	INGAN	0.48 ***	0.03	0.98 ***	0.01	0.48 ***	0.01	0.64 ***	0.02
	Ex-Sin-GAN	0.48 ***	0.02	0.98 ***	0.01	0.48 ***	0.01	0.64 ***	0.02
	PETSGAN	0.49 ***	0.02	0.98 ***	0.01	0.49 ***	0.01	0.65 ***	0.02
	PAIRGAN	0.50 ***	0.01	0.99 ***	0.01	0.50 **	0.01	0.66 **	0.01

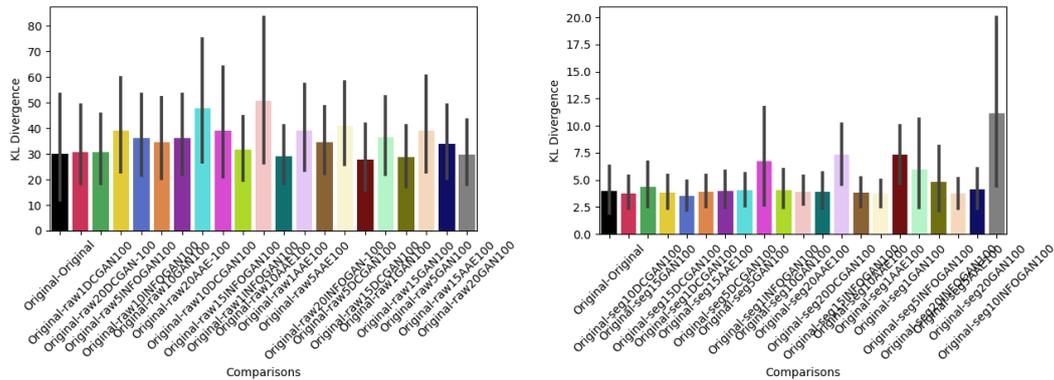


Figure 3: KL divergences on t-SNE from different generator architectures for the generation of biomedical images (left) and ground-truth (right) from few input to single input generation.

Table 6: Appendix qualitative and similitude metrics on segmented synthetic images compared to the single original segmented image. The stars correspond to *:p<0.05, **:p<0.01, ***:p<0.001.

		Blur		Mi		SSIM		Psnr	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
SI Vs All Original		1044.60	463.54	1970.92	25.46	0.61	0.09	5.43	3.26
Polyp	SinGAN	1393.56	96.55	1815.03	30.91	0.90	0.03	14.64	2.68
	CONSIN	1405.06	108.52	1808.72	20.65	0.91	0.02	15.09	2.07
	HP-VAE-GAN	1347.30	122.48	1814.70	43.28	0.90	0.05	15.00	3.16
	ONE-SHOT-GAN	1378.24	72.50	1798.90	19.78	0.92	0.02	16.07	1.92
	INGAN	1369.70	88.71	1806.53	24.77	0.91	0.02	15.56	2.32
	Ex-Sin-GAN	1359.27	106.29	1814.48	27.50	0.90	0.03	14.68	2.23
	PETSGAN	1395.62	63.91	1796.64	16.41	0.92	0.02	16.42	1.78
PAIRGAN (Ours)		1339.51	92.59	1805.07	20.99	0.91	0.02	15.47	2.02
SI Vs All Original		1314.28	143.94	-1834.47	130.65	0.46	0.13	9.45	9.95
Histological	SinGAN	357.94	27.52	-1765.31	13.14	0.09	0.02	10.31	0.51
	CONSIN	380.86	18.98	-1761.32	13.78	0.10	0.02	10.47	0.56
	HP-VAE-GAN	392.54	20.80	-1776.11	12.60	0.08	0.0	10.04	0.50
	ONE-SHOT-GAN	372.95	27.17	-1763.14	15.86	0.11	0.02	10.52	0.63
	INGAN	353.75	24.53	-1759.07	14.44	0.01	0.02	10.44	0.55
	Ex-Sin-GAN	326.89	28.06	-1747.43	15.08	0.11	0.03	10.62	0.60
	PETSGAN	333.30	27.26	-1751.00	13.28	0.12	0.02	10.76	0.57
PAIRGAN (Ours)		382.87	70.71	-1682.05	29.28	0.24	0.05	14.08	1.26
		MSE		UQM		SIFID		LPIPS	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
SI Vs All Original		19838.83	5677.25	0.57	0.10	493.37	141.04	0.30	0.05
Polyp	SinGAN	2706.06	1798.89	0.89	0.03	260.87	126.65	0.12	0.03
	CONSIN	2243.39	1056.01	0.89	0.02	227.91	75.08		
	HP-VAE-GAN	2827.56	2861.19	0.88	0.05	256.94	136.91	0.11	0.04
	ONE-SHOT-GAN	1781.11	890.11	0.90	0.02	221.83	82.55	0.10	0.02
	INGAN	2098.39	1270.77	0.90	0.03	215.38	96.59	0.10	0.03
	Ex-Sin-GAN	2538.37	1405.72	0.89	0.02	298.78	165.67	0.11	0.03
	PETSGAN	1616.76	724.83	0.91	0.02	203.15	71.88	0.09	0.02
PAIRGAN (Ours)		2064.84	1061.74	0.90	0.02	237.87	80.27	0.11	0.02
SI Vs All Original		11894.75	2939.43	0.46	0.13	200.58	52.96	0.39	0.09
Histological	SinGAN	6096.30	722.70	0.27	0.02	355.90	58.27	0.71	0.01
	CONSIN	5887.06	767.54	0.27	0.02	370.43	55.47	0.71	0.01
	HP-VAE-GAN	6486.44	749.72	0.26	0.02	366.08	63.01	0.72	0.01
	ONE-SHOT-GAN	5830.22	811.32	0.28	0.02	336.10	40.16	0.73	0.01
	INGAN	5916.38	727.33	0.27	0.02	399.93	58.59	0.70	0.01
	Ex-Sin-GAN	5696.20	781.82	0.28	0.02	373.83	43.74	0.71	0.01
	PETSGAN	5503.59	706.46	0.28	0.02	329.86	51.47	0.72	0.01
PAIRGAN (Ours)		2652.85	784.44	0.35	0.02	266.79	56.72	0.65	0.02

Table 7: Benchmark on generative architecture based upon computational comparisons.

Architecture	Time	ParametersD	ParametersG
GAN	676.30	1903875 (7.26 MB)	15492097 (59.10 MB)
DCGAN	11914.20	1903875 (7.26 MB)	433473 (1.65 MB)
INFOGAN	74987.79	1904387 (7.26 MB)	1904387 (7.26 MB)
AAE	1337.86	1904387 (7.26 MB)	142337 (556.00 KB)

Table 8: Sample from each generative architecture at 2000 epochs and for a certain number of input on histological dataset.

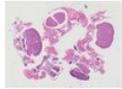
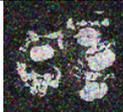
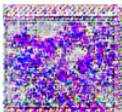
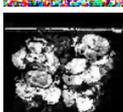
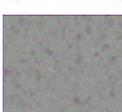
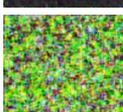
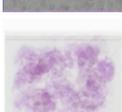
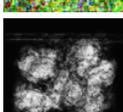
Input	Model	Raw	Seg
	original		
few input (20)	GAN		
	DCGAN		
	INFOGAN		
	AAE		
single input	GAN		
	DCGAN		
	INFOGAN		
	AAE		

Table 9: Benchmark on generative architecture based upon metric comparisons on raw images from histological dataset.

Number Input	Architecture	Blur		Mi		SSIM		Psnr	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
20	Original	1439.95	226.79	1.28	0.59	0.36	0.16	20.43	18.26
20	GAN	5179.59	2328.42	0.85	0.12	0.16	0.09	13.99	2.95
20	DCGAN	811.27	128.62	1.10	0.06	0.33	0.05	16.82	0.73
20	INFOGAN	32884.89	1097.04	0.87	0.05	0.01	0.01	4.92	0.43
20	AAE	61.75	18.92	0.70	0.09	0.35	0.04	9.45	0.29
15	GAN	4575.59	2621.84	0.94	0.11	0.22	0.08	16.03	2.38
15	DCGAN	824.38	102.77	1.07	0.07	0.31	0.07	16.73	0.77
15	INFOGAN	29011.43	21650.14	0.85	0.05	0.02	0.008	9.76	0.35
15	AAE	57.61	24.92	0.65	0.1	0.35	0.04	9.19	0.27
10	GAN	5466.76	2834.79	0.98	0.11	0.21	0.1	15.32	3.01
10	DCGAN	17726.71	456.91	0.91	0.05	0.04	0.03	8.10	0.19
10	INFOGAN	21180.05	2382.57	0.77	0.06	0.03	0.01	11.42	0.41
10	AAE	40.77	17.47	0.71	0.11	0.36	0.04	9.31	0.35
5	GAN	8928.75	4573.26	0.88	0.12	0.14	0.11	12.58	4.67
5	DCGAN	25499.19	2080.48	0.96	0.05	0.04	0.02	7.59	0.23
5	INFOGAN	12413.56	4570.68	0.75	0.10	0.06	0.04	12.06	0.41
5	AAE	29.72	22.39	0.61	0.09	0.34	0.04	8.70	0.25
1	GAN	50669.93	684.43	0.89	0.05	0.01	0.01	7.20	0.47
1	DCGAN	647.60	328.41	1.10	0.09	0.30	0.04	15.93	0.62
1	INFOGAN	4427.16	4525.23	0.52	0.22	0.17	0.10	9.63	0.53
1	AAE	130.38	5.03	1.17	0.07	0.44	0.06	19.13	0.79

Number Input	Architecture	MSE		UQM	
		Mean	Std	Mean	Std
20	Original	1480.27	89.55	0.97	0.01
20	GAN	3334.66	2782.71	0.93	0.06
20	DCGAN	1369.91	225.85	0.97	0.01
20	INFOGAN	21065.03	2088.27	0.48	0.06
20	AAE	7389.90	498.28	0.82	0.01
15	GAN	1983.19	1890.19	0.96	0.04
15	DCGAN	1400.04	229.64	0.97	0.01
15	INFOGAN	6902.15	556.23	0.88	0.01
15	AAE	7842.50	494.37	0.81	0.01
10	GAN	2627.49	3185.54	0.94	0.07
10	DCGAN	10085.92	457.35	0.76	0.01
10	INFOGAN	4705.10	450.31	0.93	0.01
10	AAE	7644.53	616.56	0.81	0.01
5	GAN	6597.83	8174.11	0.85	0.18
5	DCGAN	11351.16	600.82	0.72	0.01
5	INFOGAN	4069.32	397.58	0.93	0.012
5	AAE	8780.68	506.66	0.79	0.01
1	GAN	12454.67	1387.32	0.75	0.03
1	DCGAN	1675.49	224.75	0.96	0.01
1	INFOGAN	7129.98	790.86	0.85	0.02
1	AAE	806.59	143.83	0.98	0.01