Computer Methods and Programs in Biomedicine 240 (2023) 107695

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

ProGleason-GAN: Conditional progressive growing GAN for prostatic cancer Gleason grade patch synthesis

Alejandro Golfe[a,*], Rocío del Amor[a], Adrián Colomer[a,b], María A. Sales[c], Liria Terradez[c], Valery Naranjo[a]

[a] Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano (HUMAN-Tech), Universitat Politècnica de València, 46022, Spain
[b] ValgrAI – Valencian Graduate School and Research Network for Artificial Intelligence, Spain
[c] Anatomical Pathology Service, University Clinical Hospital of Valencia, Spain

ARTICLE INFO

Article history:
Received 17 November 2022
Revised 6 June 2023
Accepted 24 June 2023

MSC:
0000
1111

PACS:
0000
1111

Keywords:
Prostate cancer
Progressive growing GAN
Conditional GAN
Gleason grade

ABSTRACT

Background and objective: Prostate cancer is one of the most common diseases affecting men. The main diagnostic and prognostic reference tool is the Gleason scoring system. An expert pathologist assigns a Gleason grade to a sample of prostate tissue. As this process is very time-consuming, some artificial intelligence applications were developed to automatize it. The training process is often confronted with insufficient and unbalanced databases which affect the generalisability of the models. Therefore, the aim of this work is to develop a generative deep learning model capable of synthesising patches of any selected Gleason grade to perform data augmentation on unbalanced data and test the improvement of classification models.

Methodology: The methodology proposed in this work consists of a conditional Progressive Growing GAN (ProGleason-GAN) capable of synthesising prostate histopathological tissue patches by selecting the desired Gleason Grade cancer pattern in the synthetic sample. The conditional Gleason Grade information is introduced into the model through the embedding layers, so there is no need to add a term to the Wasserstein loss function. We used minibatch standard deviation and pixel normalisation to improve the performance and stability of the training process.

Results: The reality assessment of the synthetic samples was performed with the Frechet Inception Distance (FID). We obtained an FID metric of 88.85 for non-cancerous patterns, 81.86 for GG3, 49.32 for GG4 and 108.69 for GG5 after post-processing stain normalisation. In addition, a group of expert pathologists was selected to perform an external validation of the proposed framework. Finally, the application of our proposed framework improved the classification results in SICAPv2 dataset, proving its effectiveness as a data augmentation method.

Conclusions: ProGleason-GAN approach combined with a stain normalisation post-processing provides state-of-the-art results regarding Frechet's Inception Distance. This model can synthesise samples of non-cancerous patterns, GG3, GG4 or GG5. The inclusion of conditional information about the Gleason grade during the training process allows the model to select the cancerous pattern in a synthetic sample. The proposed framework can be used as a data augmentation method.

1. Introduction

Prostate cancer is the second most common cancer in men, with almost 1.4 million new cases in 2020. Once its presence has been suggested by clinical examination or laboratory tests, the main tool for diagnosis is a prostate biopsy. Tissue samples are removed with a needle, laminated, stained with haematoxylin and eosin (H&E) and stored in glass. The tissue sample is then analysed microscopically by an expert pathologist to determine the presence of cancerous patterns following the Gleason [1] classification system. This system groups the different tumour patterns

* Corresponding author.
E-mail address: algolsan@i3b.upv.es (A. Golfe).

https://doi.org/10.1016/j.cmpb.2023.107695

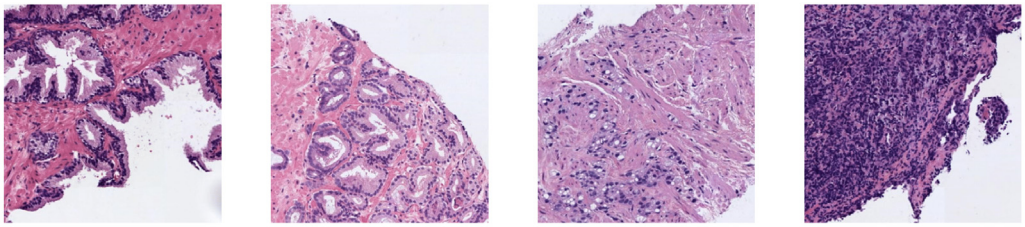

Fig. 1. Examples of patches presenting different Gleason patterns. (a) Non-cancerous well-differentiated glands; (b) Region containing GG3 atrophic dense patterns; (c) GG4 containing individual poorly-formed glands; (d) GG5 containing files of isolated cells.

into grades according to the cancer prognosis. According to Silva et al. [2], the Gleason grade system categorizes prostate cancer based on the patterns observed in the tissue. The GG3 grade includes regions with atrophic well-differentiated and dense glandular patterns. The GG4 grade contains patterns such as cribriform, ill-formed, large-fused, and papillary glandular patterns. The GG5 grade includes isolated cells, files of cells, nests of cells without lumina formation, and pseudo-rosetting patterns. Some examples of cancerous patterns belonging to different grades are shown in Fig. 1.

Pathologists detect the presence of one or more Gleason patterns by visual inspection and grade them according to the most prominent grades (e.g. a sample with two main Gleason patterns, first grade 4 followed by grade 3, would be assigned a combined grade of $4+3=7$). This combined Gleason score ranges from 6 to 10 and is assigned at the biopsy level.

In recent decades, digital pathology has become increasingly prevalent. This is a subfield of pathology that focuses on the information in digitised data and involves scanning glass slides to produce whole slide images (WSI). These high-resolution WSIs are often divided into patches of a specific resolution to obtain detailed information on all regions present in the WSI. Patch extraction allows deep learning models to process the information in them, whereas WSIs are difficult to handle due to the amount of information contained in these pyramidal images. The success of artificial intelligence and machine learning solutions combined with this type of data promotes the development of computer vision applications to automate diagnoses, prognoses and disease predictions. Deep learning (DL) approaches have shown potential in many tasks in digital pathology, as mitosis detection [3], tissue classification [4], brain tumor classification [5] and glioma grading [6].

Gleason grading is a highly time-consuming task for expert pathologists. This problem prompted the creation of artificial intelligence models capable of performing Gleason grade classification of a patch. There are many examples in the literature, e.g. Silva-Rodríguez et al. [2] and Linkon et al. [7], Arvaniti et al. [8], Bulten et al. [9]. The availability of a suitable dataset strongly conditions the performance of this models. A balanced dataset with enough samples for each class to allow learning is needed for these models to perform adequately, but this is often not possible [10]. In medical imaging and many other applications, we often deal with data where we have one sample from the minority class against hundreds of the majority. Such problems pose a challenge for predictive deep learning (DL) algorithms as most classification models have been designed under the assumption of a balanced and sufficient number of samples per class, resulting in poor classification performances, especially for minority classes. All these limitations motivate the development of a generative DL model capable of synthesising WSI patches to overcome the imbalance between classes to increase the classification model's accuracy. The very nature of the problem requires the implemented model to be able to synthesise samples conditioned to the target class (i.e. the Gleason grade to be synthesised).

In this work, we propose a conditional progressive growing GAN (ProGleason-GAN) framework able to synthesise patches of a specific Gleason grade. To the best of the author's knowledge, this is the first time in the literature that an original Progressive Growing GAN framework is modified to become a conditional GAN capable of synthesising patches of any desired Gleason grade. In the following lines, we summarize the main contributions of this paper: (i) a novel conditional Progressive Growing GAN framework for conditional image synthesis (ii) synthesis of patches of any Gleason Grade (iii) evaluation of the synthetic data by the Frechet Inception Distance (iv) validation of the model as a data augmentation method for increasing the accuracy of Gleason grading classification (v) comparison with related works in the literature (vi) systematic validation of the performance of the implemented model by a selected group of experts.

The rest of the paper is organized as follows. In Section 2, we introduce the related work present in the literature. In Section 3, we describe the database used in this work, SICAPv2, which is to the best of the author's knowledge, the largest dataset of prostate whole slide images with pixel-level annotations of the Gleason grades at patch level by expert pathologists. In Section 4, we describe the methodology followed in this research to obtain the Progressive Growing conditional GAN model. In Section 5, we show the results obtained for evaluating of the synthetic data and validating the improvement in the classification model's accuracy. Finally, Section 6 summarizes the conclusions extracted from the carried-out experiments.

2. Related work

2.1. Generative models for image synthesis

Generative Adversarial Networks (GANs) were presented by Goodfellow et al. [11] in 2014. These networks are comprised of two separate neural networks, the generator G and the discriminator D. G takes a random noise vector $z \in p_z$ as input and outputs synthetic data $G(z)$; D takes as input the output of the generator $G(z)$ and real images $x \in p_data$ to classify them as real or synthetic. The goal of training D is to maximize the probability of assigning the correct label to real $D(x)$ and synthetic data $D(G(z))$. Simultaneously, G is trained to minimize $\log(1-D(G(z)))$. Hence, G and D play a two-player min-max game with value function $V(G,D)$:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[1-\log D(G(z))] \quad (1)$$

After Goodfellow et al. introduced the first GAN framework, they left the door open to improve this architecture to synthe-

sise data conditioned to an input label. This idea prompted the development of a conditional GAN by Mirza and Osindero [12]. This Conditional GAN used a supervised approach allowing control of the class of the generated results, and had the advantage of providing better representations for a multimodal generation. GANs can be converted into conditional models by adding additional information (y) for both the generator and the discriminator. y could be auxiliary information, such as class labels or other modalities. The Eq. (1) is updated to:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x \mid y)] + \mathbb{E}_{z\sim p_z(z)}[1 - \log D(G(z \mid y))] \quad (2)$$

The previous GANs approaches showed some problems with stability and convergence. These problems encouraged the proposal of an upgraded version increasing the complexity of the network by adding CNN layers. This GAN was named *Deep Convolutional* GAN (DCGAN). After that, an extension to synthesise two dimension images was presented by Wu et al. [13]. They introduced GANs that were capable of synthesising 3D data using volumetric convolutions. This approach synthesises novel objects like chairs, tables and cars. Additionally, they proposed a method to map 2D images to 3D objects.

Later, X Chen et al. proposed in [14] a method that, rather than employing an unstructured noise vector z, decomposes the noise vector into two parts, one with incompressible noise and the other with significant structured semantic features, called c. The incompressible noise part is uncorrelated with the output, and the semantic feature part captures the structured information of the input. The authors aimed to learn a representation of the data that disentangles the high-level semantics from the incompressible factors of variation. To achieve this, the authors introduced a mutual information term in the GAN objective function, which encourages the generator to produce outputs highly dependent on the semantic features. Specifically, the mutual information term measures the information shared between the semantic feature vector and the generated output. By maximizing this term, the generator learns to create outputs that contain meaningful and structured information captured by the semantic feature vector. This approach tries to solve the following expression:

$$\min_G \max_D V_I(D,G) = V(D,G) - \lambda I(c; G(z,c)) \quad (3)$$

where $V(D,G)$ is the objective function presented in the original GAN approach, I is the mutual information, $G(z,c)$ is the synthetic sample, and λ is a regularization parameter. The objective is to maximize the mutual information between c and $G(z,c)$ maximizing $I(c; G(z,c))$. The difference between InfoGAN with conditional GANs is that the latent code c is not known, it is learned during the training process.

All proposed methods created synthetic samples from random input noise but could not do the reverse operation. To address this, Donahue et al. proposed the BiGAN model [15]. This was the first time in the literature where the synthetic samples are mapped to their latent vector representation to determine which features better represent the characteristics in the generated samples. An architecture present in the literature that has been employed for style transfer is CycleGAN [16]. Style transfer refers to transforming an image from one domain to another while preserving its content. CycleGAN, consisting of two generators and two discriminators, enables style transfer without needing paired image datasets. This innovation allows for applying different styles to images, even when paired style-reference images are not available. The Pix2Pix architecture [17] is also noteworthy for style transfer but requires paired images for training. Moreover, Karras et al. [18] proposed the first approach by introducing progressive growth training. This framework was named Progressive Growing GAN (ProGAN) and represented the central core of this paper. This fact motivates an in-depth explanation in the following sections of this paper, in which we also remark on the contributions of our proposed framework.

Furthermore, a prominent architecture in the literature for style transfer is StyleGAN [19]. This architecture is designed explicitly for generating high-quality images with fine-grained control over the style and appearance of the output. It has been successfully utilized in various applications, including artistic style transfer and domain adaptation. StyleGAN introduces a novel mapping network that enables the disentanglement of style and structure, allowing for the transfer of different visual characteristics across domains.

2.2. Generative models for histological image synthesis

There are numerous applications of GANs in histological image analysis [20]. Specifically, significant applications include stain normalisation, stain and domain adaptation, segmentation with supervised models, synthesis enabling weakly supervised and unsupervised learning, and data generation and augmentation for classification purposes.

Stain normalisation involves mapping an original image to a normalised domain, reducing variability. In the study by Zhou et al. [21], they introduced a stain normalisation technique by leveraging CycleGAN tailored for this purpose. Another approach to stain normalisation, as described in Zanjani et al. [22], involved utilizing InfoGAN, where the latent information was substituted with the lightness channel of the source image. In this scenario, the generator was trained using mutual information loss to learn the structured colour space, enabling the transformation of colour from the original domain to the normalised domain. CycleGAN has also been utilized for stain and domain adaptation. The distinction from the previously mentioned method lies in its focus on domain shifting between different staining techniques. Some examples in this field include the study by Xu et al. [23], where they adapted the CycleGAN architecture to handle samples with other stainings, such as H&E and IHC. Furthermore, in the approach presented in Swiderska-Chadaj et al. [24], the creation of histological image patches is proposed by cropping and combining different patches with a smooth blending of the seams using a CycleGAN. This method is referred to as multi-patch blending. Another architecture employed for this purpose is Pix2Pix [17]. In contrast to the approaches mentioned above that utilize unpaired data, this architecture uses paired data. An example of the application of this architecture can be found in Rana et al. [25], where Pix2Pix is employed to obtain unstained images from H&E images.

The aforementioned Pix2Pix network has proven to be an alternative to conventional fully convolutional methods [26]. In this study, the architecture was adapted to the field of histology for the basal membrane segmentation of microinvasive cervix carcinoma. Another application of this architecture for segmentation can be found in Cheng et al. [27]. In this work, masks are generated from generated points, which are then translated into synthetic tissue samples.

Regarding the use of GANs for data generation and augmentation, which is the main focus of this article, several noteworthy works in the literature are mentioned below. Wei et al. [28] adapted the CycleGAN architecture for data augmentation. Instead of performing domain adaptation, they trained their network to switch between normal and abnormal domains. In this way, they successfully implemented a generative model that, given samples from one class, could generate their equivalents in the other class.

Examples of using the DCGAN architecture in histological image synthesis can be found in Xue et al. [29], Krause et al. [30]. In the work by Y. Xue et al., they introduced modifications to the original

DCGAN architecture to enable conditional synthesis of cervical cancer samples based on their class. This framework was designed as a data augmentation method to enhance their classification models' accuracy for different cervical cancer classes. They reported an increase in accuracy from 66.3% to 71.7%. The generated samples had a resolution of 256x128; however, objective metrics assessing the quality of the generated data were not provided, with the focus being solely on the improved accuracy of sample classification. Furthermore, J. Krause et al. proposed a novel DCGAN approach for synthesizing histopathological images of colorectal cancer. This architecture incorporated an embedding layer to concatenate label information with the input data, enabling conditional synthesis by the DCGAN. The approach adopted in Oyelade et al. [31] shares similarities with the DCGAN methodology but with a specific focus on abnormalities associated with breast images. The training was carried out on a category-based basis, emphasizing the detection and characterization of breast abnormalities.

By leveraging the DCGAN framework, the authors in Karimi et al. [32] presented an application of conditional GAN and DCGAN for synthesizing prostate patches of size 192^2 based on the Gleason score. Specifically, the conditional GAN network was able to be trained on the entire dataset, while their DCGAN approach lacked conditional capability, leading them to train a separate model for each class. Another example in the literature combining the ideas of conditional GAN and DCGAN is the ProstateGAN approach proposed in Hu et al. [33]. This architecture was used to synthesize focal prostate diffusion images of size 32^2.

In [34], a ProGAN approach that could synthesise brain tumour histopathological images was presented. In this study, two distinct ProGAN models were trained separately, each focusing on one of the two classes in their dataset. The authors demonstrate that training a dedicated ProGAN model for each class and incorporating synthetic data into the dataset resulted in a 5% increase in their classification models' accuracy. Another work using this architecture was presented by Teramoto et al. [35]. In this case, a ProGAN architecture was used to synthesise lung cancer histological images. They aimed to improve the accuracy of their classification models to classify their samples into benign or malignant. More specifically, they reported a 4.3% increase in the accuracy of their models. It is worth noting that the last two frameworks presented did not have the capacity for conditional synthesis. In [36], they proposed a multi-scale conditional GAN for high- resolution, large-scale histopathology image generation and segmentation. This model is composed of a hierarchical arrangement of GAN structures, with each level dedicated to generating and segmenting images at a distinct scale.

Moreover, a novel conditional deep learning architecture based in StyleGAN [19] and BigGAN [37] was proposed in Quiros et al. [38] in which colorectal and breast cancer samples were synthesised. In recent years, modern architectures such as transformers have been employed in histological imaging for image synthesis. For instance, in MedViTGAN [39], the use of a conditional GAN without convolutions based on transformers is proposed as a method for data augmentation.

Therefore, this work proposes an approach that combines the progressive training technique with conditional synthesis for histopathological image synthesis, specifically for generating $10\times$ magnification patches containing the specified cancerous pattern based on the input condition related to the Gleason scale. Moreover, the evaluation using the FID metric and a publicly available dataset with pixel-level annotations of the Gleason grade for the patches enables a more objective assessment for future research and comparisons. As discussed in this article, this dataset exhibits sufficient diversity of patterns in local structures that the model can learn. This fact offers an advantage over other methods proposed in the literature as it can learn more complex and diverse patterns. Additionally, the performance improvement of classification models is evaluated using the proposed method as a data augmentation technique. Furthermore, to enhance the present study's completeness and validity, a group of expert pathologists was selected to conduct a study that directly evaluates the quality of representation for each cancerous pattern in the synthetic samples, as described in this article. The combination of all these aspects represents an innovative approach within the current literature, highlighting the unique contribution of our proposed method to the field.

3. Materials: SICAPv2 database

The database used in this study was presented in Silva-Rodríguez et al. [2] and it is publicly available at SICAPv2 dataset. This is the most extensive public collection of prostate H&E biopsies with patch-level annotations of Gleason grades.

To the best of the author's knowledge, there exist five main databases containing prostate cancer tissue images. The Cancer Genome Atlas project released the largest database of up to 720 prostate biopsy slides [40]. However, the absence of annotations for Gleason grades at both the local and biopsy levels restricts the utility of these data [2]. Another database shared by Arvaniti et al. [8] provides pixel-level annotations of Gleason patterns for 886 small regions of slides (cores of TMAs). Unfortunately, these cores do not adequately represent the diverse patterns found in local structures of prostate cancer and benign lesions, thus lacking clinical relevance for slide-level Gleason score diagnosis. Similar limitations exist in the database presented at the Gleason19 challenge in the MICCAI 2019 conference [41], which includes 331 annotated cores by different pathologists and the dataset used in Ing et al. [42] comprising 625 isolated patches. Finally, another large dataset was presented in the PANDA challenge [43]; however, its Gleason score labels were at the biopsy level, making it not directly applicable to this study.

SICAPv2 includes 155 biopsies from 95 patients who signed informed consent. Tissue samples were sliced, laminated, stained with Hematoxylin and Eosin (H&E) and digitised using the Ventana iScan Coreo scanner at 40x magnification to obtain WSI. Expert pathologists analysed the slides obtained at Hospital Clínico of Valencia and assigned a combined Gleason score per biopsy. In cases where the Gleason grade of a sample was uncertain, experts set the label by consensus to avoid inter-observer variability. To handle large WSI, we down-sampled them to patches of $10\times$ resolution of size 512^2 and overlap of 50% between them. Patches with less than 20% of tissue without cancerous patterns annotated by the pathologists were discarded. After this procedure, SICAPv2 contained 4417 non-cancerous patches, 2222 labelled as GG3, 4494 as GG4, and 948 as GG5. A summary of the database description is presented in Table 1.

Table 1
SICAPv2 description. Amount of WSI with their respective biopsy-level and number of patches for each Gleason Grade.

WSI				
Non cancerous	GG3	GG4	GG5	Total
37	60	97	16	182
Patches				
Non cancerous	GG3	GG4	GG5	Total
4417	2222	4494	948	12,081

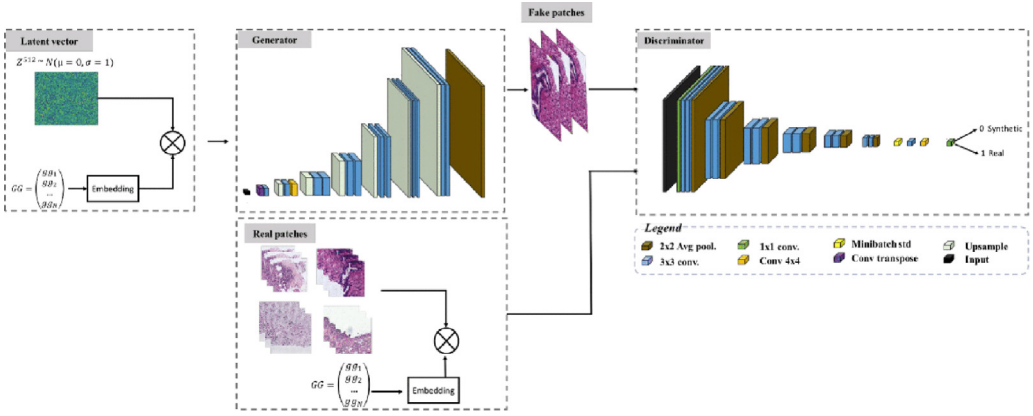

Fig. 2. Overview of the proposed prostate histopathological image synthesis with conditional progressive growing generative adversarial network.

4. Methodology

The proposed framework is based on a conditional progressive growing GAN (ProGleason-GAN) able to synthesise prostate histology patches of any Gleason grade. The workflow, which is composed of a generator θ^g and discriminator θ^d, is presented in Fig. 2. The details of each component are given in Appendix A.

4.1. CGAN for prostate image synthesis

The methodological core of this work is based on a conditional generative adversarial network. In this sense, the generator model aims to produce synthetic prostate histology patches containing the cancerous pattern associated with the Gleason grade. Formally, we denote the random input noise as $Z = \{z_1, \ldots, z_i, \ldots, z_M\}$, where z_i is the ith instance obtained from a normal distribution $\mathcal{N}(\mu = 0, \sigma = 1)$ and M is the total number of generated samples. We note as N the total number of patches in the dataset, and for convenience, we give M the same value as N. Additionally, the generator is provided with the specified Gleason grade. We denote it as $GG = \{gg_1, \ldots, gg_i, \ldots, gg_M\}$, where gg_i represents the ith Gleason grade associated with the ith noise instance. The values of the gg_i instances range from 0 to 3 for non-cancerous, GG3, GG4 and GG5, respectively. Therefore, the synthetic patch generation is defined as follows:

$$I = f(Z, GG; \theta^g) \tag{4}$$

where $I \in \mathbb{Z}^{m \times n \times 3}$ represents all the generated synthetic prostate histology patches. Here, m represents the height, n represents the width, and the number 3 refers to the three RGB channels. Additionally, θ^g denotes the model weights.

The discriminator model aims to classify input patches as real (1) or fake (0). Let us define $X = \{x_1, \ldots, x_i, \ldots, x_N\}$, where x_i represents the ith instance of real prostate histology patches. The input to the discriminator is $B = X \cup I$. The objective is to predict ($\hat{Y}$) for each instance, which can be defined as follows:

$$\hat{Y}_b = f(B, GG; \theta^d) \tag{5}$$

where θ^d denotes the discriminator model weights.

In the following subsections, we explain the minibatch standard deviation and pixel normalisation methods, which are used to enrich the variety of learning on the training data and guarantee their stability. Finally, we introduce the loss function used in this framework.

Minibatch standard deviation. GANs naturally tend to learn only a subset of the training dataset. To solve this problem, we use a "minibatch discrimination" [44]. We compute the standard deviation across the feature maps in the minibatch, encouraging the generated images to show similar statistics to training images. The proposal does not have any learnable parameters or hyperparameters. First, we compute the standard deviation for each feature in each spatial location across the minibatch. We reduce all the statistics computed to a single value by averaging over all the features and spatial locations. Then, we replicate this value and concatenate it to all spacial locations across the minibatch, creating an additional feature map. We can place this layer anywhere in the discriminator, but it performs better if inserted towards the end.

Pixel normalisation. Sometimes the magnitudes of the generated values in the generator and the discriminator spiral out of control due to their competition. To solve these problems, we normalise the feature vector in each pixel to unit length in the generator after each convolution layer. We use a variant of the "local response normalisation" proposed in Hinton et al. [45]. The expression of the pixel normalisation is shown in Eq. (6), where L represents the number of feature maps, $(a^j_{x,y})$ the pixel to be normalised, and ϵ the error to avoid zero values.

$$b_{x,y} = \sqrt{\frac{1}{L} \cdot \sum_{j=0}^{L-1} (a^j_{x,y})^2 + \epsilon} \tag{6}$$

Loss function. To optimize the proposed model, we used the Wasserstein GAN with Gradient Penalty (WGAN-GP) loss function [46]. First, we set the learning rate η, the c value that sets the maximum oscillation range of the gradients to $[-c, c]$ and the batch size bs. A batch of real $x^{(i)}$ and synthetic data $G(z^{(i)})$ is sampled. The loss function of the discriminator is implemented as Eq. (7):

$$loss_D = \frac{1}{bs} \cdot \sum_{i=1}^{bs} (D(x^{(i)})) - \frac{1}{bs} \cdot \sum_{i=1}^{bs} (D(G(z^{(i)}))) \tag{7}$$

After computing the gradients, the weights of the discriminator are updated. Then, we define the loss function of the generator as follows:

$$loss_G = \frac{1}{bs} \cdot \sum_{i=1}^{bs} (D(G(z^{(i)}))) \tag{8}$$

WGAN-GP improves the training of GAN as it aims to minimize the distance between two probabilistic distributions, which are the distribution of the real data and the synthetic one.

4.2. ProGleason-GAN

Due to the complexity of the problem, the CGAN framework cannot learn to synthesise patches of the target resolution 256^2. It is necessary to train the model progressively at lower resolutions

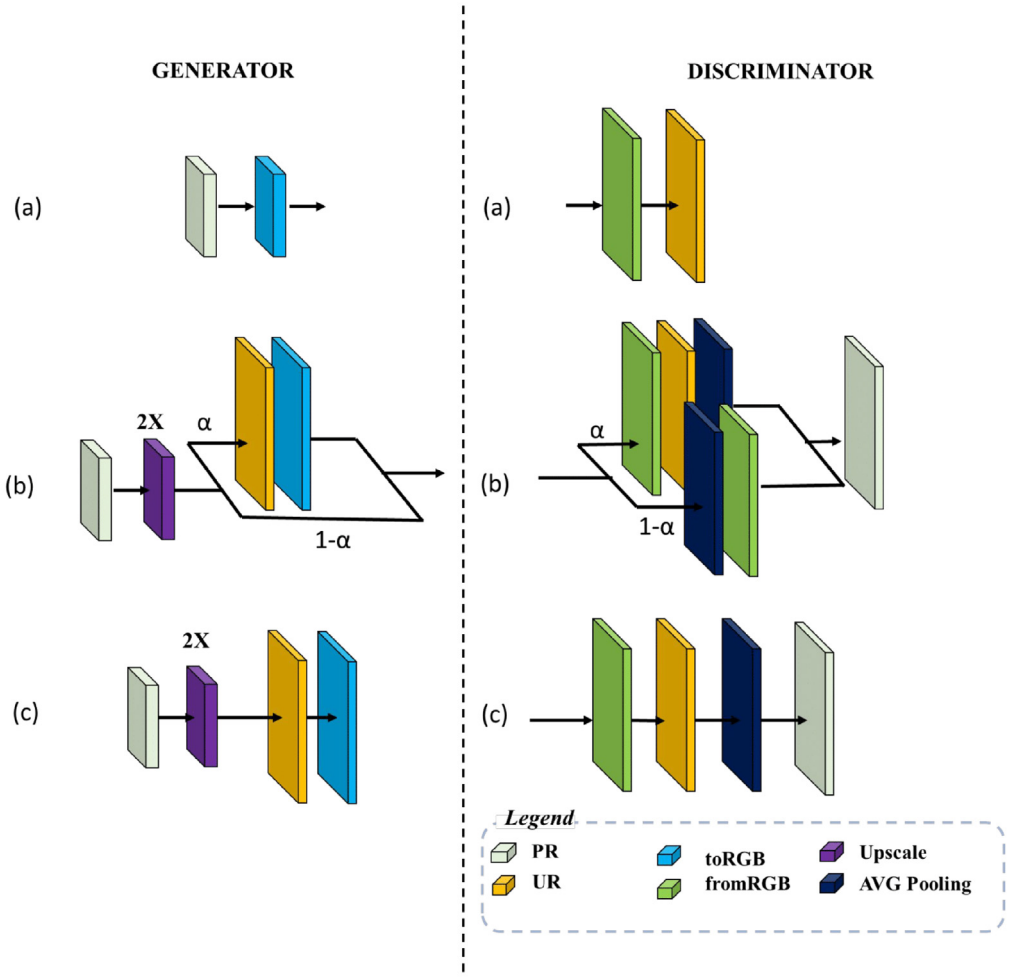

Fig. 3. Fade-in new resolution layers to the discriminator and generator.

to solve this problem. Progressively increasing the size of the network has several benefits. As we increase the resolution little by little, we ask the network to learn a more straightforward question than learning how to generate the target resolution from scratch. We start training with 4^2 resolution patches and then progressively increase the resolution to 256^2, adding layers to the networks as described in Fig. 3. This way of training allows the model to learn high-level features of the image distribution first and then progressively increase the complexity of the details instead of learning the whole information at once.

The original *Progressive Growing GAN* architecture proposed by Karras et al. [19] is modified to introduce conditional information about the Gleason grade. This modification allows the trained network to synthesise patches containing the specified Gleason grade cancerous pattern. The generator and the discriminator receive this information during the training process. In the generator case, the Gleason grade is given after an embedding layer in charge of converting the value to a fixed-length vector of defined size. This vector is concatenated to the input noise. Moreover, the discriminator receives the information about the Gleason grade, and the embedding layer transforms it into a fixed-length vector concatenated to the input data as an additional channel. In this work, we perform the conditional synthesis without introducing any additional term to the loss function. The introduction of the embedding information about the Gleason grade encourages the generator and the discriminator to learn the features and the difference between each class.

The generator and the discriminator grow both simultaneously. All layers in both networks remain trainable during the training process. We add new resolution layers, fading them as illustrated in Fig. 3. This method prevents previous smaller-resolution layers from suffering a sudden shock. Fading in the new layers provides more stability to the training process when we double the resolution of the generator (G) and discriminator (D). Figure 3 shows how it was increased from the prior resolution (PR) in (a) to the upscaled resolution (UR) in (c) in the case of the generator, while the reverse is done in the discriminator. Transition (b) shows how the new resolution layers are treated as residual blocks whose weight α increases linearly from 0 to 1. This value rises progressively, starting from a small value until it reaches 1 in the last training epoch. In this way, the new layer is introduced gradually to prevent destabilization of the training process. ToRGB represents a layer that transforms feature vectors to RGB space and does the reverse operation from RGB. We use an upscale method with the nearest neighbour algorithm to increase the resolution in the generator and average pooling to do the opposite in the discriminator, both by a factor of two.

We start training both generator (G) and discriminator (D) at the lowest resolution, 4^2 increasing the resolution by a factor of 2 until we reach the target resolution, 256^2.

Table 2
SICAPv2 split. Number of patches selected for the training and test subsets.

Train partition				
Non cancerous	GG3	GG4	GG5	Total
3773	1829	3641	716	9959
Test partition				
Non cancerous	GG3	GG4	GG5	Total
644	393	853	232	2122

Table 4
Weighted FID results for all the frameworks considered in this work: CGAN, ProGAN and ProGleason-GAN. Additionally, we provide the metrics obtained after the stain normalisation process.

Method	FID
CGAN	160.55
CGAN + Stain Norm	207.22
ProGAN	126.51
ProGAN + Stain Norm	86.13
ProGleason-GAN	120.14
ProGleason-GAN + Stain Norm	**77.85**

5. Experiments and results

This section shows the experiments carried out to validate the proposed framework. First, the results of the proposed method (ProGleason-GAN) are shown and compared with the original ProGAN. In addition, the ProGleason-GAN's performance is validated using a staining normalisation method to reduce the variability of the different stains. Then, the synthetic sample quality and its ability to represent the different Gleason grades are validated by a group of experts. Finally, the effectiveness of the proposed model as a data augmentation strategy is quantitatively evaluated.

5.1. Experimental setting

Database partitioning. We used the partition given in Silva-Rodríguez et al. [2] to split the database in the training and test sets. To avoid model overestimation, the splitting was performed at patient level (see Table 2).

Implementation. All the validated experiments were implemented using Pytorch version 1.9.1 and Python 3.7. Experiments were conducted on the NVIDIA DGXA100 system. The code is publicly available on ProGleason-GAN GitHub repository

Model hyper-parameters. The optimal hyper-parameters combination was achieved by training the models during 100 epochs, using Adam optimizer, with a 0 value for β_1 and 0.99 for β_2, a learning rate of 0.001 and the WGAN-GP as loss function. In the case of batch size, it was set to 64 for resolutions from 4^2 to 128^2 and 32 for 256^2. The optimal size of the generator input was 512.

Evaluation. We used the Frechet Inception Distance (FID) [47] to evaluate the proposed model. This metric allows assessing the difference between two multidimensional Gaussian distributions. In this case, features corresponding to synthetic and real patches were extracted using the Inception V3 model trained on the ImageNet dataset [48]. We denote the feature distribution of synthetic and real patches as $\mathcal{N}(\mu, C)$ and $\mathcal{N}(\mu_w, C_w)$, respectively. The FID expression is shown in the following equation:

$$FID = \| \mu - \mu_w \|^2 + Tr(C + C_w - 2(C \cdot C_w)^{\frac{1}{2}}) \quad (9)$$

Note that FID $\in [0, +\infty]$, being 0 the optimal value. We obtained the FID metric for each Gleason grade and a weighted average to obtain the global FID (see Table 3 for the class weight). The purpose of using these weights is based on the fact that each Gleason grade has a different number of samples and, therefore, different representations in the dataset.

In addition, we calculated the area under the ROC curve (AUC) of the classification carried out by the expert group, and precision, F1-Score and accuracy metrics for the data augmentation strategy validation.

5.2. Ablation experiments

Quantitative results. Table 4 shows the weighted FID results for all the frameworks considered in this work: CGAN, ProGAN and ProGleason-GAN. Progressive training using the ProGAN framework provides a significant performance improvement compared to the CGAN approach. However, as previously mentioned, the ProGAN architecture has no conditional synthesis capacity. The proposed conditional progressive framework (ProGleason-GAN) provides the best performance regarding the FID metric evaluation. After the staining normalisation process proposed in Macenko et al. [49], the best results are obtained by the proposed method, particularly, a 35.2% improvement in the weighted FID metric. The effect of stain normalisation has been evaluated on all proposed methods to assess its impact on the FID metric results. It is remarkable that, in the case of CGAN, it leads to a degradation of the results. This can be attributed to the low quality of the generated samples and their dissimilarity to the original ones, which may result in the stain normalisation excessively altering these samples and further deviating them from the originals. Conversely, for ProGAN and ProGleason-GAN, stain normalisation improves the results. ProGleason-GAN achieves the best outcomes, even when the effect of stain normalisation is eliminated.

Table 5 shows the FID results obtained for the different frameworks at each Gleason grade. As the original ProGAN method has no conditional synthesis capacity, its results were not included in Table 5.

These results show the underperformance of the CGAN architecture compared to the proposed model, as the complexity of the network is not adequate to learn the features of the input images. As expected, since non-cancerous and GG4 have more patches for learning, ProGleason-GAN obtains the best results for these classes.

Qualitative results. To qualitatively evaluate the proposed method, we provide images synthesised by the CGAN, ProGAN and

Table 3
SICAPv2 weight distribution for the test subset.

Class weights	
Gleason grade	Test
Non-cancerous	0.3035
G3	0.1852
G4	0.4020
G5	0.1093

Table 5
FID for conditional experiments. NC for non cancerous patterns and GG3, GG4, GG5 for each Gleason grade.

Test				
	NC	GG3	GG4	GG5
CGAN	198.45	172.51	152.2	165.7
CGAN + Stain Norm	219.43	217.43	191.18	214.98
ProGleason-GAN	92.14	128.53	127.86	155.35
ProGleason-GAN + Stain Norm	**88.85**	**81.86**	**59.32**	**108.69**

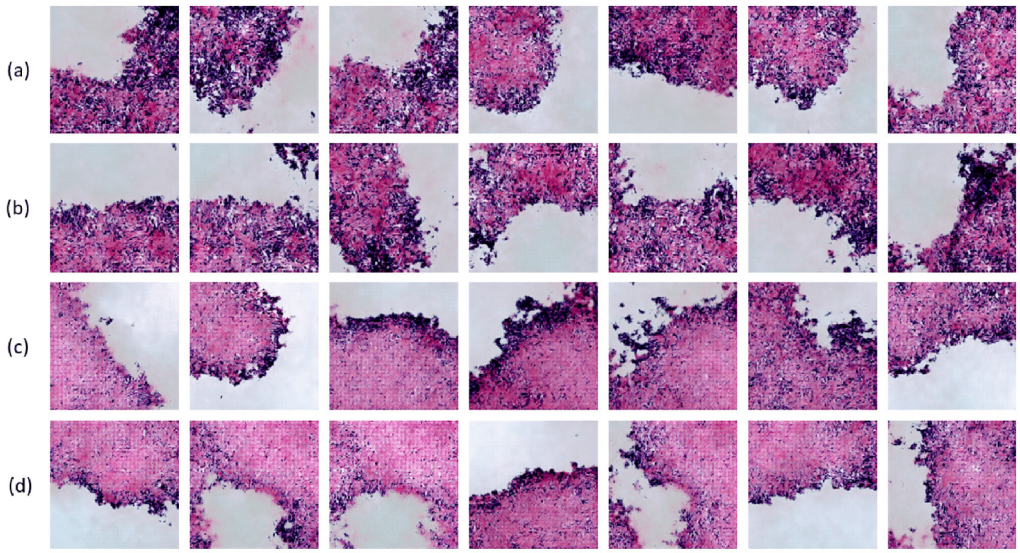

Fig. 4. Synthetic patches generated with CGAN framework. (a) Non cancerous; (b) GG3; (c) GG4; (d) GG5.

ProGleason-GAN frameworks. Figure 4 shows some patches generated with the CGAN approach. This framework cannot learn the morphology and features of each Gleason grade. It captures color properties and spatial distribution but cannot produce complex structures such as glands or nuclei.

Some patches generated with the ProGAN are shown in Fig. 5. This framework is able to generate samples with the real patch morphology and distribution, but it can not reproduce with detail the cancerous patterns present in SICAPv2. The structure morphology is unclear and seems to be a combination of different structures in all Gleason grades but not representing each.

Figure 6 shows some synthesised examples by ProGleason-GAN. By introducing conditional information in a progressive training, it is observed that samples of different Gleason grades are different and share features with samples of the same class. Therefore, the proposed model can identify each class's intrinsic features.

Figure 7 shows some samples generated for each Gleason grade by ProGleason-GAN after the stain normalisation post-processing. The generated samples and the real data showed more homogeneity. The non-cancerous images show well-differentiated glands. GG3 images show how the density of glands increases and irregularities appear. Furthermore, GG4 represents a high-grade and poorly differentiated carcinoma. Finally, GG5 shows the least differentiated cancerous pattern, where a high density of disorganized nuclei characterizes chaotic tissue behavior. Therefore, synthetic images show the same patterns as the real ones.

ProGleason-GAN is capable of synthesizing prostate tissue patches that are sufficiently realistic. One of the advantages provided by our model is the absence of a specific term in the loss function for conditional synthesis. This simpler approach reduces complexity in the network and promotes stability in the learning process during training. Additionally, employing the FID metric for

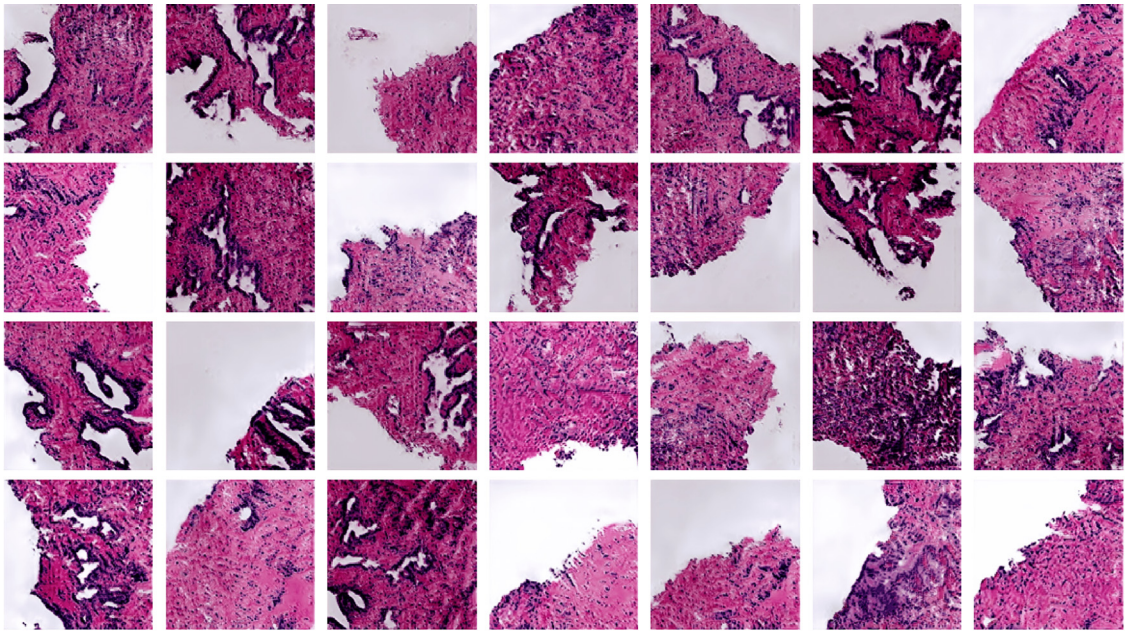

Fig. 5. Synthetic patches generated with the original ProGAN framework.

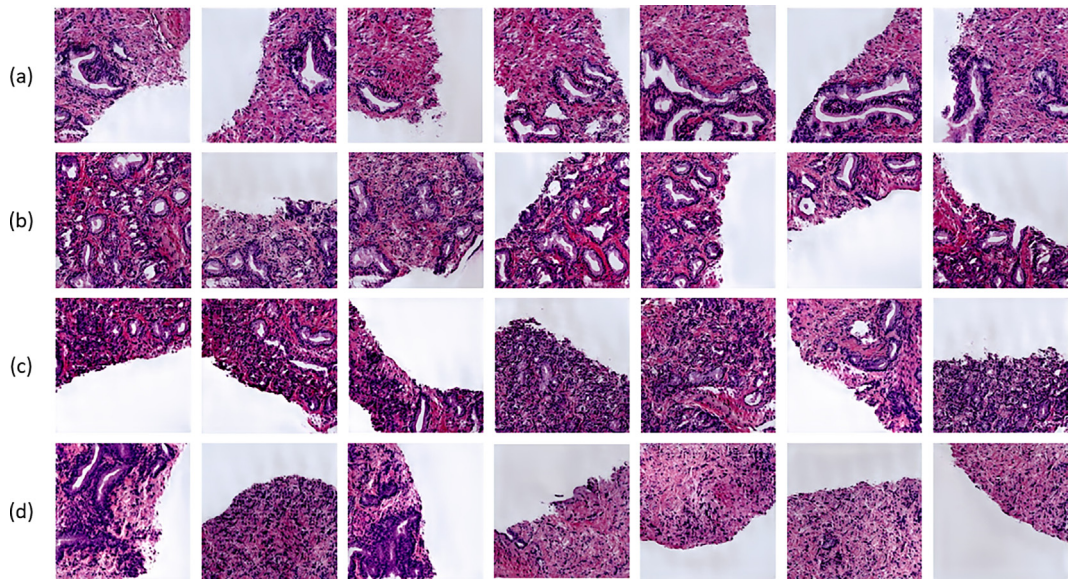

Fig. 6. Synthetic patches generated with ProGleason-GAN. (a) Non cancerous; (b) GG3; (c) GG4; (d) GG5.

evaluation enables future comparisons with other GAN architectures trained on the same dataset for this task. A direct comparison with the existing literature would be inaccurate since the employed metric is conditioned on the dataset used. Another notable contribution of this approach is that it represents a step forward in the on-demand synthesis of complete WSI, enabling not only the synthesis of the constituent patches but also the selection of the cancerous pattern contained within them.

Our method outperforms existing approaches in several vital aspects. In [32], they focus on synthesizing patches of size 192^2. In contrast, our method achieves a higher resolution of 256^2, resulting in more detailed and visually appealing synthetic images. Additionally, the study by Karimi et al. [32] does not provide a comprehensive evaluation metric such as FID, which is essential for objectively assessing the quality of the generated images. In contrast, our method incorporates FID evaluation, allowing for more objective comparison and validation of the results. Furthermore, their DCGAN approach requires training a separate model for each class, whereas our method is capable of handling all data in a more unified manner. This simplifies the training process and enables our model to capture and synthesize the distinctive features of various classes effectively. Regarding the work presented in Hu et al. [33], their focus lies in synthesizing prostate diffusion images at a resolution of 32×32, which is significantly lower and of a different typology compared to the approach presented in this study. The complexity of this type of images is also considerably lower than the images addressed in our work. In [36], they employ the FID metric for evaluating their results; however, their baseline model is not the standard InceptionV3, preventing direct comparison. Furthermore, their dataset is smaller, consisting of approximately 1500 patches. Additionally, their dataset does not have participant-level partitioning, making it uncertain whether patches from the same patient are used in both the training and testing phases.

5.3. External validation protocol

To demonstrate the usefulness of the proposed method, a panel of experts validated the quality of the synthetic images. Experts

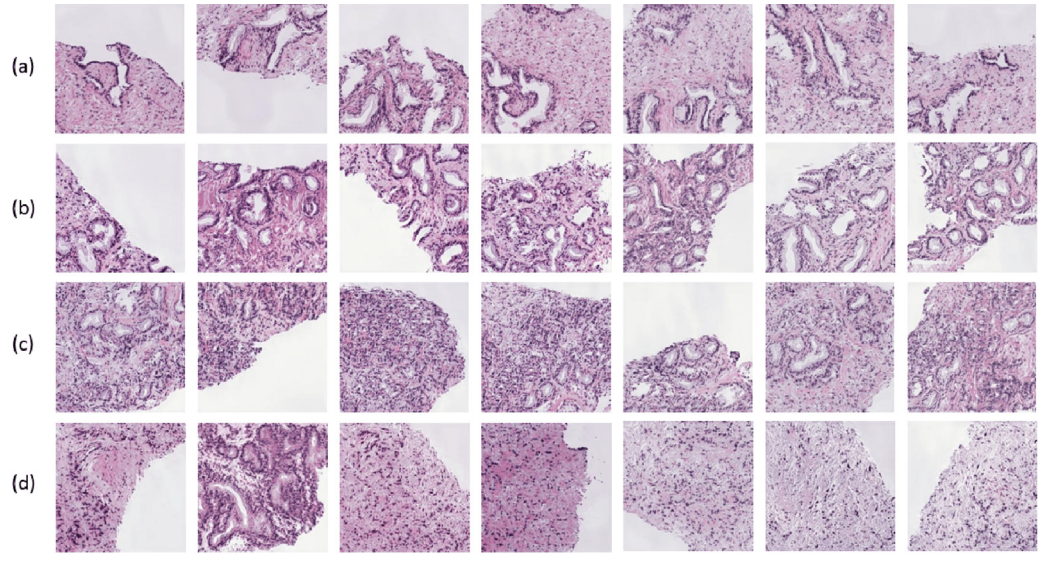

Fig. 7. Synthetic patches generated with ProGleason-GAN framework and stain normalisation. (a) Non cancerous; (b) GG3; (c) GG4; (d) GG5.

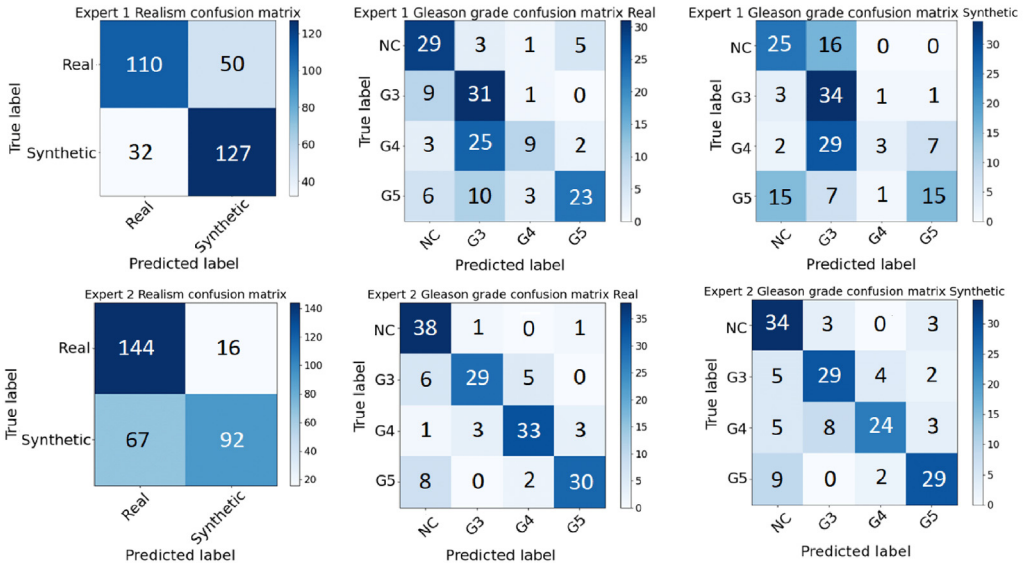

Fig. 8. Confusion matrix for expert pathologist classification. (Left) Classification for real and synthetic samples; (Middle) Gleason grade classification for real samples; (Right) Gleason grade classification for synthetic ones.

were asked to identify whether a sample was real or synthetic and to establish its Gleason grade. In total, 320 samples (160 real and 160 synthetic) were analysed. In addition, all Gleason grades were equally represented, resulting in 40 samples for each grade and image type (real and synthetic).

Figure 8 shows the results obtained by the expert panel. Expert pathologist incorrectly identified on average 20.6% of the real samples as synthetic. As for synthetic images, approximately 31% were considered to be real. Regarding the Gleason grading, GG4 was better detected in synthetic samples, and approximately the 75% of the GG5 samples were identified in synthetic images, one of the most critical cancerous patterns to be appropriately recognized due to their advanced tumoral stage.

Finally, we tested if there was a statistically significant difference in establishing gleason grade on synthetic or real images. For this purpose, the area under the ROC curve (AUC) metric was used. Table 6 shows the AUC metric for classifying synthetic (S) and real (R) patches for each expert pathologist. In addition, the p-value (with $\alpha = 0.05$) is provided.

Table 6 show no statistically significant difference between performing the grading with real or synthetic images. This fact demonstrates that the implemented model correctly learned to distinguish and represent the different Gleason grades. Regarding GG4 detection by the Expert 2, a p-value close to the 95% confidence interval limit is obtained. This could be due to the same patch may contain different cancer patterns according to the Gleason scale. This fact introduces inaccuracies during the learning process, as the most prevalent Gleason grade of a sample was considered as a label.

Table 6
AUC metric for gleason grading with real (AUC-R) and synthetic (AUC-S) images and p-value from the statistical inference study in the expert group.

	Expert 1			Expert 2		
	AUC-R	AUC-S	p-value	AUC-R	AUC-S	p-value
NC	0.8078	0.7201	0.1886	0.9125	0.8458	0.1964
GG3	0.7183	0.7192	0.9904	0.8458	0.8166	0.6239
GG4	0.5947	0.5281	0.3767	0.8833	0.775	0.0667
GG5	0.7441	0.6643	0.2625	0.8583	0.8291	0.6132

Table 7
Results for the patch-level Gleason grading in the test set for the model proposed in Silva-Rodríguez et al. [2] with the original SICAPv2 dataset (S) and upsampling with our proposed data augmentation method (S + P). The metrics presented are precision, F1-Score, computed per class, and global accuracy.

	Precision		F1-S		ACC	
	S	S + P	S	S + P	S	S + P
NC	0.8081	**0.8376**	0.8348	**0.8473**	–	–
GG3	0.5096	**0.5529**	0.4908	**0.5981**	–	–
GG4	0.6394	**0.7087**	0.6667	**0.6934**	–	–
GG5	0.5714	**0.6284**	0.4301	**0.5542**	–	–
Avg	0.6321	**0.6819**	0.6056	**0.6733**	0.6673	**0.7078**

5.4. Data augmentation strategy validation

This section shows the validation of the proposed method as a data augmentation strategy. For this purpose, we compare the classification model used in Silva-Rodríguez et al. [2] trained with SICAPv2 and SICAPv2 augmented with the proposed model. Specifically, the minority classes (GG3 and GG5) were augmented by 20%, increasing 366 GG3 samples and 144 GG5 samples. The results obtained are shown in Table 7. The proposed model significantly improves the classification model performance for all classes. This fact supports the effectiveness and validity of the proposed work as a patch synthesis method and data augmentation strategy.

6. Conclusions

In this study, we propose a conditional Progressive Growing GAN framework to synthesize prostate tissue patches with any Gleason Grade. The proposed framework obtained a weighted FID metric for all Gleason grades of 77.85, compared to the 160.55 and 120.14 achieved by the CGAN and ProGAN, respectively. To asses the quality of the synthetic samples, a group of expert pathologists performed an external validation. The statistical study determined

no significant difference in establishing the Gleason grade with synthetic or real samples. Additionally, we pretrained a classification model using the synthesized images and SICAPv2 dataset. The proposed method improved the classification accuracy by 4.05% compared to the network fine-tuned only with SICAPv2. These findings confirmed the effectiveness of the use of ProGleason-GAN-generated images.

Concerning the limitations of this study, the resolution and generated image size, as well as the amount of training data, may also pose challenges to the model's applicability in clinical settings. For instance, while the model's resolution was adequate for this study, using higher resolutions may be necessary for clinical environments, where images with more detail are required for diagnosis and treatment planning. It should be noted that in this work, the resolution of the synthesized images was constrained to 256^2 due to hardware restrictions. Achieving higher resolutions, such as 512^2 or 1024^2, is possible by incrementing the model complexity of the proposed approach (and, of course, the hardware resources).

Moreover, the limited size of the training database, coupled with the inherent variability of clinical data, may reduce the model's generalizability to different clinical scenarios. It is worth emphasizing that obtaining databases with pixel-level annotation of prostate patches based on the Gleason score pattern can be a challenging task.

In conclusion, while this study provides valuable insights into the potential utility of deep learning in prostate cancer detection, it is essential to acknowledge the identified limitations, including those of the model's applicability in clinical settings. Future research should strive to address these limitations and evaluate the model's performance in real-world clinical scenarios. In future directions, our research will be centred on the synthetic generation of complete Whole Slide Images (WSIs) leveraging the findings and insights from this study.

Funding

This work has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation, under grant agreement no. 860627 (CLARIFY), the Spanish Ministry of Economy and Competitiveness through project PID2019-105142RB-C21 (AI4SKIN) and GVA through projects PROMETEO/2019/109 and INNEST/2021/321 (SAMUEL). The work of Adrián Colomer has been supported by the ValgrAI – Valencian Graduate School and Research Network for Artificial Intelligence & Generalitat Valenciana and Universitat Politècnica de València (PAID-PD-22). Rocío del Amor has been supported by the Spanish Government under FPU Grant (FPU20/05263).

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

Acknowledgements

We gratefully acknowledge the support from the Generalitat Valenciana (GVA) with the donation of the DGX A100 used for this work, action co-financed by the European Union through the Operational Program of the European Regional Development Fund of the Comunitat Valenciana 2014–2020 (IDIFEDER/2020/030).

Appendix A. Detailed generator and discriminator architectures

Tables A.8 and A.9.

Table A.8
Architecture of the generator for 256^2 resolution.

Generator	Activation	Output shape
Latent vector + Embedding (Gleason Grade)	–	$(512 + 512) \times 1 \times 1$
Pixel Norm	–	$1024 \times 1 \times 1$
Conv transpose 4×4	Leaky ReLU	$512 \times 4 \times 4$
Conv 3×3	Leaky ReLU	$512 \times 4 \times 4$
Upsample	–	$512 \times 8 \times 8$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 8 \times 8$
Conv 4×4 + Pixel Norm	Leaky ReLU	$512 \times 8 \times 8$
Upsample	–	$512 \times 16 \times 16$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 16 \times 16$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 16 \times 16$
Upsample	–	$512 \times 32 \times 32$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 32 \times 32$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 32 \times 32$
Upsample	–	$512 \times 64 \times 64$
Conv 3×3 + Pixel Norm	Leaky ReLU	$256 \times 64 \times 64$
Conv 3×3 + Pixel Norm	Leaky ReLU	$256 \times 64 \times 64$
Upsample	–	$256 \times 128 \times 128$
Conv 3×3 + Pixel Norm	Leaky ReLU	$128 \times 128 \times 128$
Conv 3×3 + Pixel Norm	Leaky ReLU	$128 \times 128 \times 128$
Upsample	–	$128 \times 256 \times 256$
Conv 3×3 + Pixel Norm	Leaky ReLU	$64 \times 256 \times 256$
Conv 3×3 + Pixel Norm	Leaky ReLU	$64 \times 256 \times 256$
Conv 1×1	Leaky ReLU	$3 \times 256 \times 256$

Table A.9
Architecture of the discriminator for 256^2 resolution.

Discriminator	Activation	Output shape
Input image + Embedding (Gleason Grade)	–	$(3+1) \times 256 \times 256$
Conv 1×1 + Pixel Norm	Leaky ReLU	$16 \times 256 \times 256$
Conv 3×3 + Pixel Norm	Leaky ReLU	$16 \times 256 \times 256$
Conv 3×3 + Pixel Norm	Leaky ReLU	$32 \times 256 \times 256$
Average Pooling	–	$32 \times 128 \times 128$
Conv 3×3 + Pixel Norm	Leaky ReLU	$32 \times 128 \times 128$
Conv 3×3 + Pixel Norm	Leaky ReLU	$64 \times 128 \times 128$
Average Pooling	–	$64 \times 64 \times 64$
Conv 3×3 + Pixel Norm	Leaky ReLU	$64 \times 64 \times 64$
Conv 3×3 + Pixel Norm	Leaky ReLU	$128 \times 64 \times 64$
Average Pooling	–	$128 \times 32 \times 32$
Conv 3×3 + Pixel Norm	Leaky ReLU	$128 \times 32 \times 32$
Conv 3×3 + Pixel Norm	Leaky ReLU	$256 \times 32 \times 32$
Average Pooling	–	$256 \times 16 \times 16$
Conv 3×3 + Pixel Norm	Leaky ReLU	$256 \times 16 \times 16$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 16 \times 16$
Average Pooling	–	$512 \times 8 \times 8$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 8 \times 8$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 8 \times 8$
Average Pooling	–	$512 \times 8 \times 8$
Minibatch standard deviation	–	$(512+1) \times 8 \times 8$
Conv 3×3 + Pixel Norm	Leaky ReLU	$512 \times 8 \times 8$
Conv 4×4 + Pixel Norm	Leaky ReLU	$512 \times 1 \times 1$
Conv 1×1	–	$1 \times 1 \times 1$

References

[1] D.F. Gleason, Histologic grading of prostate cancer: a perspective, Hum. Pathol. 23 (3) (1992) 273–279.
[2] J. Silva-Rodríguez, A. Colomer, M.A. Sales, R. Molina, V. Naranjo, Going deeper through the Gleason scoring scale: an automatic end-to-end system for histology prostate grading and cribriform pattern detection, Comput. Methods. Programs Biomed. 195 (2020) 105637.
[3] M. Veta, P.J. Van Diest, S.M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A.B. Larsen, J.S. Vestergaard, A.B. Dahl, et al., Assessment of algorithms for mitosis detection in breast cancer histopathology images, Med. Image Anal. 20 (1) (2015) 237–248.
[4] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, A. Madabhushi, Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks, in: Medical Imaging 2014: Digital Pathology, vol. 9041, SPIE, 2014, p. 904103.
[5] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images, IEEE Trans. Med. Imaging 35 (1) (2015) 119–130.
[6] M.G. Ertosun, D.L. Rubin, Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks, in: AMIA Annual Symposium Proceedings, vol. 2015, American Medical Informatics Association, 2015, p. 1899.
[7] A.H.M. Linkon, M.M. Labib, T. Hasan, M. Hossain, Marium-E-Jannat, Deep learning in prostate cancer diagnosis and Gleason grading in histopathology images: an extensive study, Inform. Med. Unlocked 24 (2021) 100582, doi:10.1016/j.imu.2021.100582.
[8] E. Arvaniti, K.S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P.J. Wild, J.H. Rueschoff, M. Claassen, Automated Gleason grading of prostate cancer tissue microarrays via deep learning, Sci. Rep. 8 (1) (2018) 1–11.
[9] W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, J. van der Laak, C. Hulsbergen-van de Kaa, G. Litjens, Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study, Lancet Oncol. 21 (2) (2020) 233–241, doi:10.1016/S1470-2045(19)30739-9.
[10] K. Alomar, H.I. Aysel, X. Cai, Data augmentation in classification and segmentation: a survey and new strategies, J. Imaging 9 (2) (2023) 46.
[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[12] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784(2014).
[13] J. Wu, C. Zhang, T. Xue, B. Freeman, J. Tenenbaum, Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling, Adv. Neural Inf. Process. Syst. 29 (2016).
[14] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: interpretable representation learning by information maximizing generative adversarial nets, Adv. Neural Inf. Process. Syst. 29 (2016).
[15] J. Donahue, P. Krähenbühl, T. Darrell, Adversarial feature learning, arXiv preprint arXiv:1605.09782(2016).
[16] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
[17] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[18] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation, arXiv preprint arXiv:1710.10196(2017).
[19] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4401–4410.
[20] M.E. Tschuchnig, G.J. Oostingh, M. Gadermayr, Generative adversarial networks in digital pathology: a survey on trends and future potential, Patterns 1 (6) (2020) 100089.
[21] N. Zhou, D. Cai, X. Han, J. Yao, Enhanced cycle-consistent generative adversarial network for color normalization of H&E stained images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22, Springer, 2019, pp. 694–702.
[22] F.G. Zanjani, S. Zinger, B.E. Bejnordi, J.A. van der Laak, P.H. de With, Stain normalization of histopathology images using generative adversarial networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 573–577.
[23] Z. Xu, X. Huang, C.F. Moro, B. Bozóky, Q. Zhang, GAN-based virtual re-staining: a promising solution for whole slide image analysis, arXiv preprint arXiv:1901.04059(2019).
[24] Z. Swiderska-Chadaj, E. Stoelinga, A. Gertych, F. Ciompi, Multi-patch blending improves lung cancer growth pattern segmentation in whole-slide images, in: 2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE), IEEE, 2020, pp. 1–4.
[25] A. Rana, G. Yauney, A. Lowe, P. Shah, Computational histological staining and destaining of prostate core biopsy RGB images with generative adversarial neural networks, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018, pp. 828–834.
[26] D. Wang, C. Gu, K. Wu, X. Guan, Adversarial neural networks for basal membrane segmentation of microinvasive cervix carcinoma in histopathology images, in: 2017 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 2, IEEE, 2017, pp. 385–389.
[27] J. Cheng, Z. Wang, Z. Liu, Z. Feng, H. Wang, X. Pan, Deep adversarial image synthesis for nuclei segmentation of histopathology image, in: 2021 2nd Asia Conference on Computers and Communications (ACCC), IEEE, 2021, pp. 63–68.
[28] J. Wei, A. Suriawinata, L. Vaickus, B. Ren, X. Liu, J. Wei, S. Hassanpour, Generative image translation for data augmentation in colorectal histopathology images, Proc. Mach. Learn. Res. 116 (2019) 10.
[29] Y. Xue, Q. Zhou, J. Ye, L.R. Long, S. Antani, C. Cornwell, Z. Xue, X. Huang, Synthetic augmentation and feature-based filtering for improved cervical histopathology image classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 387–396.
[30] J. Krause, H.I. Grabsch, M. Kloor, M. Jendrusch, A. Echle, R.D. Buelow, P. Boor, T. Luedde, T.J. Brinker, C. Trautwein, et al., Deep learning detects genetic alterations in cancer histology generated by adversarial networks, J. Pathol. 254 (1) (2021) 70–79.
[31] O.N. Oyelade, A.E. Ezugwu, M.S. Almutairi, A.K. Saha, L. Abualigah, H. Chiroma,

A generative adversarial network for synthetization of regions of interest based on digital mammograms, Sci. Rep. 12 (1) (2022) 6166.
[32] D. Karimi, G. Nir, L. Fazli, P.C. Black, L. Goldenberg, S.E. Salcudean, Deep learning-based Gleason grading of prostate cancer from histopathology images–role of multiscale decision aggregation and data augmentation, IEEE J. Biomed. Health Inform. 24 (5) (2019) 1413–1426.
[33] X. Hu, A.G. Chung, P. Fieguth, F. Khalvati, M.A. Haider, A. Wong, ProstateGAN: mitigating data bias via prostate diffusion imaging synthesis with generative adversarial networks, arXiv preprint arXiv:1811.05817(2018).
[34] S. Liu, Z. Shah, A. Sav, C. Russo, S. Berkovsky, Y. Qian, E. Coiera, A. Di Ieva, Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning, Sci. Rep. 10 (1) (2020) 1–11.
[35] A. Teramoto, T. Tsukamoto, A. Yamada, Y. Kiriyama, K. Imaizumi, K. Saito, H. Fujita, Deep learning approach to classification of lung cytological images: two-step training using actual and synthesized images by progressive growing of generative adversarial networks, PLoS One 15 (3) (2020) e0229951.
[36] W. Li, J. Li, J. Polson, Z. Wang, W. Speier, C. Arnold, High resolution histopathology image generation and segmentation through adversarial training, Med. Image Anal. 75 (2022) 102251.
[37] A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv preprint arXiv:1809.11096(2018).
[38] A.C. Quiros, R. Murray-Smith, K. Yuan, PathologyGAN: learning deep representations of cancer tissue, arXiv preprint arXiv:1907.02644(2019).
[39] M. Li, C. Li, P. Hobson, T. Jennings, B.C. Lovell, MedvitGAN: end-to-end conditional GAN for histopathology image augmentation with vision transformers, in: 2022 26th International Conference on Pattern Recognition (ICPR), IEEE, 2022, pp. 4406–4413.
[40] R. Burton, M.A. Jensen, A. Kahn, T. Pihl, D. Pot, Y. Wan, D.A. Levine, Tissue Source Site, The cancer genome atlas pan-cancer analysis project, Nat. Genet. 45 (10) (2013) 1113–1120.
[41] Gleason 2019 dataset, 2019, (Online). https://gleason2019.grand-challenge.org/Home/.
[42] N. Ing, Z. Ma, J. Li, H. Salemi, C. Arnold, B.S. Knudsen, A. Gertych, Semantic segmentation for prostate cancer grading by convolutional neural networks, in: Medical Imaging 2018: Digital Pathology, vol. 10581, SPIE, 2018, pp. 343–355.
[43] W. Bulten, K. Kartasalo, P.-H.C. Chen, P. Ström, H. Pinckaers, K. Nagpal, Y. Cai, D.F. Steiner, H. van Boven, R. Vink, et al., Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the panda challenge, Nat. Med. 28 (1) (2022) 154–163.
[44] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, X. Chen, Improved techniques for training GANs, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, vol. 29, Curran Associates, Inc., 2016. https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf
[45] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580(2012).
[46] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of Wasserstein GANs, Adv. Neural Inf. Process. Syst. 30 (2017).
[47] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local Nash equilibrium, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf
[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[49] M. Macenko, M. Niethammer, J.S. Marron, D. Borland, J.T. Woosley, X. Guan, C. Schmitt, N.E. Thomas, A method for normalizing histology slides for quantitative analysis, in: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2009, pp. 1107–1110.