SAFRON: Stitching Across the Frontier for Generating Colorectal Cancer Histology Images

Srijay Deshpande, Fayyaz Minhas, Simon Graham, Member, IEEE, Nasir Rajpoot, Senior Member, IEEE

Abstract—Synthetic images can be used for the development and evaluation of deep learning algorithms in the context of limited availability of annotations. In the field of computational pathology where histology images are large and visual context is crucial, synthesis of large tissue images via generative modeling is a challenging task due to memory and computing constraints hindering the generation of large images. To address this challenge, we propose a novel framework named as SAFRON to construct realistic large tissue image tiles from ground truth annotations while preserving morphological features and with minimal boundary artifacts at the seams. To this end, we train the proposed SAFRON framework based on conditional generative adversarial networks on large tissue image tiles from the Colorectal Adenocarcinoma Gland (CRAG) and DigestPath datasets. We demonstrate that our model can generate high quality and realistic image tiles of arbitrary large size after training it on relatively small image patches. We also show that training on synthetic data generated by SAFRON can significantly boost the performance of a standard algorithm for gland segmentation of colorectal cancer tissue images. Sample high resolution images generated using SAFRON are available at the URL: https://warwick.ac.uk/TIALab/SAFRON

Index Terms—Computational Pathology, Generative Adversarial Networks, Image Synthesis, Deep Learning, Annotated Data Generation

I. INTRODUCTION

AUTOMATED analysis of histology whole-slide images (WSIs) has received a great deal of attention in the field of medical image analysis in recent years. Several deep learning approaches for problems in the area of computational pathology have been proposed. These include tumor segmentation [1], [2], segmentation of glands [3], cancer grading [4], [5] and nuclei detection and classification [6]–[8]. Digital WSIs are typically quite large in size, potentially containing several billions of pixels at the highest magnification. Therefore, development of efficient machine learning algorithms for histology slides remains a major challenge, due to limitations in processing capacity and memory storage.

Deep learning models are data hungry in nature, and consequently developed neural models for image analysis usually require large and high quality annotated datasets. This problem is exacerbated in medical image analysis, where data annotation needs to be done by an expert pathologist. To overcome the difficulty of curating large datasets, generative modelling of synthetic images has become an active area of research in recent years and has resulted in the development of numerous models for synthetic tissue image generation in the area of computational pathology (CPath). For instance, Kovacheva et al. [9] presented a parameterised generative model to produce synthetic histology image data, where the user has control over the input parameters like cancer grade and cellularity. More recently, motivated by the ability of Generative adversarial networks (GANs) [10] to use adversarial training to generate realistic natural images with high perceptual quality, conditional Generative Adversarial Networks (GANs) have been proposed to synthesise histology images [11], [12]. Quiros et al. proposed Pathology-GAN [13] to generate high quality cancer tissue images. Although these models can successfully generate high quality realistic images, they can only generate smaller sized images due to limited memory and processing capacity. Large image tiles are useful because they provide additional context, which can assist a diagnosis in computational pathology [4], [14].

In this paper, we propose SAFRON (Stitching Across the Frontier), which is a framework for generating annotated high quality synthetic histology images. Specifically, the framework generates large colorectal tissue images based on provided input tissue component masks by methodically stitching together generated tissue regions within a large tissue tile. In order to generate realistic images with a high perceptual quality, our framework utilises adversarial training. To the best of our knowledge, this is the first framework that can generate histology images of large arbitrary sizes. We evaluate our model on the CRAG [3] and DigestPath [15] datasets and show that the constructed images are capable of generating morphological features like glands, goblet cells and stromal regions. We also highlight the potential of using SAFRON...
Fig. 1: Concept diagram of SAFRON along with its application

to generate a large synthetic dataset, which can assist with training of gland segmentation models. The concept diagram of our proposed framework is illustrated in Figure 1.

The main contributions of the proposed work are:

1) We present a computationally cheap and memory efficient framework, termed as SAFRON, that can generate tissue tiles of much larger sizes than the ones used for training. The SAFRON framework shows the ability of seamless generation of large tiles preserving homogeneity and edge crossing continuities between the adjacent patches.

2) We show that SAFRON can generate very high dimensional synthetic colorectal tissue tiles of dimensions greater than $8000 \times 5000$ pixels. As far as we are aware, it is the first framework that can generate such large tissue images.

3) Our proposed framework has the ability to generate an unlimited amount of synthetic data. As an example, we show that augmenting training data for gland segmentation with the synthetic data generated by SAFRON shows significant improvement in terms of segmentation accuracy.

4) Our qualitative and quantitative evaluation on CRAG [3] and DigestPath [15] datasets shows that the synthetic images are constructed with precision preserving morphological characteristics, and can be useful for training and evaluation of gland segmentation algorithms.

The remainder of this paper is organised as follows. In the next section, we present the details of the SAFRON framework, explanation of its each components. We perform detailed evaluation of the proposed framework on the CRAG [3] and DigestPath [15] datasets. Later, we evaluate our model with both qualitative and quantitative metrics and show the usage of annotated synthetic data for gland segmentation. Finally, we conclude with future directions for this work.

II. THE PROPOSED SAFRON FRAMEWORK

The construction of large-size synthetic images is computationally demanding due to the high computational complexity and memory requirements of image generation neural networks such as variational auto-encoders [16] and standard generative adversarial networks [10]. In this work, we propose a novel method which can generate large tissue images of arbitrary sizes conditioned on input tissue component masks. The proposed method, called Stitching Across the FRonTier Generative Adversarial Network (SAFRON), aims to provide annotated synthetic data which can then be used for the training and evaluation of computational histopathology algorithms. The proposed framework attempts to generate small tissue image patches based on local tissue components such as glands, stroma and background specified as an input mask and stitches the generated patches in a seamless manner, thus, overcoming computational limitations.

A. Datasets

For training and performance evaluation of the proposed SAFRON framework, we require annotated real image data in terms of tissue component specification for conditioning the generation of synthetic images. In this work, we consider two datasets for our experiments: CRAG [3], [17] and Digestpath [15].

The CRAG dataset has been widely used for the development of numerous gland segmentation methods [3], [18], [19]. It [3] consists of 213 colorectal histopathology cancerous tissue images with variable cancer grades and a size of $1512 \times 1516$ pixels at a resolution of 0.55m/pixel (20 objective magnification). Each image in the dataset belongs to one of three classes based on structure and morphology of glands: normal/healthy (no deformation of glands), low-grade cancer (slightly deformed glands), and high-grade cancer (highly deformed glands). Each image in the dataset is also associated with a gland segmentation mask which specifies glandular, non-glandular (stromal tissue) and background (based on raw pixel thresholding) regions. The input mask can be seen on the left part of Fig. 1 where the glands are shown in green, the stroma in red and the background in blue. This input segmentation mask is used as a tissue component mask to condition the generation of synthetic images through the proposed framework. In this paper, we have focused on generation of images with normal grades only. The dataset contains 48 large colon tissue cancerous images with normal grades.}

1 https://warwick.ac.uk/TIALab/data/mildnet/.
grade of which we use 39 (the CRAG train set) to extract training data and 9 (CRAG test set) to extract testing data.

We also use the DigestPath [15] (DP) dataset to assess the performance of our algorithm. The dataset is collected from the DigestPath2019 challenge. Similar to the CRAG dataset, it also contains tissue images with corresponding glandular masks. However, the image tiles are generally larger with an average size of around $5000 \times 5000$ pixels. This dataset originally contains annotations for malignant lesions only. In order to obtain a tissue segmentation mask that shows glandular, stromal and background regions, we used a semi-automatic approach. For this purpose, we first trained trained a gland segmentation model [3] on the GlaS [20], [21] and CRAG [3] datasets and obtained gland segmentation masks for images in the DP dataset which were manually refined. We have used a total of 46 images from this dataset for our experiments: 20 images (DP train set) to extract training data and 26 images (DP test set) for extraction of testing data.

B. Architecture

The proposed framework is based on generative adversarial neural networks. A conventional generative adversarial network (GAN) [10] consists of two components, a generator and a discriminator, which are trained simultaneously such that the generator learns to produce realistic images from an underlying image distribution whereas the discriminator attempts to discriminate between real and generated images. On successful training on real images, the generator is used for generating realistic images which are not part of the training set. Conditional generative adversarial networks [22] are capable of generating images controlled by some ground-truth input.

Figure 2 presents an overview of the proposed framework. The input to the framework is a tissue component mask which specifies different tissue regions to be generated by the proposed network. The tissue component mask can be obtained from an annotated dataset as described above or constructed by a user. Similar to existing methods for histology image generation based on conditional adversarial networks, the proposed framework also uses a generator network to generate small tissue image patches from a patch of the input tissue component mask. The patches are then then stitched to generate large image tiles. However, the major challenge here is to preserve global coherence and edge crossing continuity between adjacent patches so that the generated tile appears homogeneous and seamless while staying within image size limitations imposed by memory and computational processing constraints. In SAFRON, the size of the input tissue component mask patch is kept larger ($296 \times 296$) than that of the output tissue image patch ($256 \times 256$) to encourage context aware generation without seams at patch borders. Furthermore, in contrast to existing methods in which both generator and discriminator networks work at patch level, the generator network in SAFRON generates ($256 \times 256$) patches whereas the discriminator network is trained over ($728 \times 728$) tile-level images after stitching. Since the SAFRON discriminator treats the generated tile as a single large image, this enforces global coherence with minimal computational overhead. Additionally, adjacent patches are constructed with a small overlap which helps to reduce boundary artifacts. Furthermore, since the generator generates small ($256 \times 256$) image patches, it has a low memory and computational processing fingerprint.

Since the patch-level generator is responsible for generating patches of tissue tiles from corresponding input mask patches and stitching of adjacent patches is independent of tile size, the proposed framework can be used to construct large histology images of arbitrary sizes. Intuitively, the framework learns to generate local regions and stitch them in a way that the generated tile exhibits a seamless appearance without boundary artifacts between adjacent regions. Below, we discuss different
components of the proposed framework in detail.

C. Tissue Components Masks Generation

A tissue component mask is used to specify different tissue regions as input to the proposed framework. For training and performance evaluation, we use tissue segmentation masks in CRAG and DP datasets as discussed in the Dataset section above. However, for some experiments, we use two additional approaches to construct tissue component masks: i) Random Shape Generation, and, ii) TheCoT Spatial Model of Tumour Heterogeneity in Colorectal Adenocarcinoma Tissue by Kovacheva et al. [9].

In random shape generation, we first place elliptical objects of variable sizes as glands on a blank tissue component mask of desired size. A total of 3 to 7 such objects are added at random spatial locations in a (100 × 100) pixel area. These ‘glandular’ objects are colored in green to discriminate them from stromal (‘red’) and background (‘blue’) regions which are added in the remaining area with random color filling using a binomial distribution with probabilities of 0.9 and 0.1, respectively.

TheCoT constructs synthetic images along with the gland segmentation mask based on input parameters like cancer grade, cellularity, cell overlap ratio, image resolution and objective level. The gland segmentation masks generated by this model can be used as tissue component masks in SAFRON with stromal and background regions added in a similar manner as used for random shape generation. In comparison to random shape generation above, this allows us to generate tissue component masks in a more parameter-controlled manner.

Sample tissue component masks from these approaches are shown in Figure 3

D. Patch Generation

We denote a given input component mask and corresponding real histology image as $X$ and $Y$, respectively, which can be modelled as ordered sets of patches, i.e., $X = \{x_{r,c}\}$ and $Y = \{y_{r,c}\}$ with each patch parameterized by its center grid coordinates $(r,c)$ in the corresponding image. It is important to note that the size of an input component mask patch is kept larger $(296 \times 296)$ than the size of tissue patch $(256 \times 256)$ to encourage context aware image generation. Furthermore, patches are generated with an overlap of 20 pixels between adjacent patches which allows for seamless stitching.

The proposed framework generates a tissue image patch $y'$ from a corresponding tissue component mask patch $x$ using a generator neural network $G$, i.e., $y' = G(x; \theta_G)$, where $\theta_G$ denotes the trainable weights of the generator.

E. Stitching

The patches generated by the SAFRON generator network $Y' = \{y'_{r,c}\}$ for a given input tissue component mask are stitched based on the spatial coordinates of their corresponding tissue component mask patches. Specifically, pixels in overlapping regions between adjacent patches are spatially averaged in the stitching process. The results of the stitching operation on one of the test image is shown in figure 4. It clearly shows that the stitched image does not have any seam or block artifacts after training.

F. Generator and Discriminator Networks

We adapt the pix2pix generator and PatchGAN discriminator architectures given in [23] for our framework. Similar to U-Net [24], the generator, $G(x; \theta_G)$ has a U-shaped structure with two components: encoder and decoder. The architecture of the generator is shown in Fig. 5. The encoder takes a tissue component mask patch as input and constructs a low dimensional representation. The decoder then generates a tissue image patch corresponding to the input component mask patch based on encoder outputs.

The encoder is made up of a series of encoding blocks, where each block performs the following operation: pass the input through a convolutional layer followed by batch normalization and finally output after applying leaky-relu activation. The decoder consists of a series of decoding blocks: each block upsamples the input to a higher dimension by first passing the input through a deconvolutional layer, followed by batch normalization and leaky-relu activation. The generator utilises skip-connections between symmetric encoding and decoding blocks in the generator. Skip-connections give the generator flexibility to bypass the encoding part to subsequent layers and enable consideration of low level features from earlier encoding blocks in the generator.

The discriminator network in SAFRON, $D(X,Z; \theta_D)$ with trainable parameters $\theta_D$, assigns a degree of realism to a generated or a real tile $Z$ through its association with the corresponding tissue component mask $X$. As shown in [3], the discriminator has a series of encoding blocks which take a $(728 \times 728)$ image tile $Z$ and the corresponding mask $X$ as input and produce a $20 \times 20$ matrix that signifies the degree of realism of corresponding regions in $Z$. These elements of this $20 \times 20$ matrix are averaged to generate a single decision score from the discriminator.

G. Adversarial Training and Inference

For training, we use tissue component masks and their corresponding real image tiles of size $(728 \times 728)$ which are obtained from images in a training dataset. The generator and discriminator networks are trained in an adversarial manner such that the generator tries to generate realistic images whereas the discriminator aims to determine if an image is real or generated from the generator. The generator $G(x; \theta_G)$ uses $(296 \times 296)$-sized tissue component mask patches $x \in X$ to generate corresponding $(256 \times 256)$ patches $y' = G(x; \theta_G)$ of the tissue image which are stitched to generate a $(728 \times 728)$ tile image. Without introducing further notation, we denote the stitched tile image corresponding to an input tissue component mask $X$ as $Y' = G(X; \theta_G)$. The discriminator network $D(X,Z; \theta_D)$ produces a realism score for a tile $X$ which then used in Adversarial training. Adversarial training ensures that the generator learns to generate seamless images with global morphological coherence.
Fig. 3: Variety of tissue component masks used in this work to generate realistic tissue tiles. From left, first two component masks are collected from existing datasets CRAG and DigestPath (DP), third component mask is generated randomly whereas the fourth one is constructed using Kovacheva’s method [9].

Fig. 4: Stitching operation: four spatially adjacent constructed patches from the generator are stitched (averaging the overlapped portion), combined region is shown in the middle figure which looks homogeneous and seamless. The merged portion at the middle of region is zoomed and shown in the rightmost figure.

The complete framework with trainable parameters \(\{\theta_G, \theta_D\}\) is trained by minimizing a loss function with the following two components:

**Reconstruction Loss:** This loss component captures tile reconstruction error between original and generated tiles after stitching. Specifically, we use the expected value of the L1 loss between actual and generated tile in the training dataset as given as below:

\[
L_R(\theta_G) = E_{X,Y \sim \text{data}}(X,Y) \| Y - G(X; \theta_G) \|_1
\]  

**Adversarial Loss:** The generator and discriminator networks are trained in an adversarial manner with a cross-entropy based loss function which forces the generator to generate realistic images and the discriminator to classify between real and generated images:

\[
L_{\text{adv}}(\theta_G; \theta_D) = E_{X,Y \sim \text{data}}(X,Y)[\log D(X,Y; \theta_D)] + E_{X \sim \text{p}_X(X)}[\log(1 - D(X, G(X, \theta_G); \theta_D))]
\]

The overall learning problem can then be expressed as a the following adversarial optimization problem based on the linear combination of adversarial and reconstruction losses:

\[
\min_{\theta_G} \max_{\theta_D} \lambda_1 L_R(\theta_G) + \lambda_2 L_{\text{adv}}(\theta_G; \theta_D)
\]  

The hyper-parameters, \(\lambda_1\) and \(\lambda_2\), control the relative contributions and scaling of reconstruction and adversarial losses and their values are set to 1 and 100, respectively, through cross-validation based tuning.

To train the SAFRON framework on the CRAG dataset, we extract 975 square tiles of size 728 \(\times\) 728 with a stride of 200. For the purpose of testing, we extract 144 square tiles of size 728 \(\times\) 728, 81 tiles of size 964 \(\times\) 964, 81 tiles of size 1200 \(\times\) 1200 and 36 tiles of size 1436 \(\times\) 1436 from the set of CRAG test set. Similarly in order to train the model on the DigestPath dataset, we extract 971 images of size 728 \(\times\) 728 from 20 very large images and set aside 26 for testing. The network is trained for 100 epochs with Adam optimization and an initial learning rate \(10^{-4}\), initial momentum 0.5 and batch size of one tile with multiple patches.

It is important to note that the proposed framework is trained to generate globally consistent patches given an input tissue component mask. Consequently, at inference time, we can use the trained SAFRON model to generate a large image of arbitrary size as the stitching mechanism is independent of the tile size that was used for training.

### III. EXPERIMENTATION AND RESULTS

In this section, we present visual and quantitative results of the quality of generated images. We demonstrate how synthetic images can be used for training and performance assessment of gland segmentation algorithms. We also show how SAFRON generated data can be used to augment limited training datasets and improve data efficiency of gland segmentation methods.

#### A. Visual Assessment

Fig. 6a shows generated tiles from the DigestPath test dataset after SAFRON is trained on DigestPath train dataset. The first row shows a generated tile of size 8743 \(\times\) 5631 along with a zoomed region from the same tile. For this purpose, the real tissue component mask of the image was used as input and this allows us to compare the generated image with the corresponding real image. We observe that, shapes, morphological characteristics and glandular appearances are

[^3]: https://warwick.ac.uk/TIALab/SAFRON
preserved in the generated tile and the generated tile resembles the corresponding real image very closely. The quality of stitching operation is also shown in figure 6a where we can see seamless edge-crossing continuities. Though constructed tiles appear seamless and homogeneous, it can be noticed in the high resolution image that surface epithelium is not constructed with high fidelity.

The visual results on a representative image from the CRAG test dataset are shown in Fig. 7. For this purpose, SAFRON was trained on the CRAG train set. Similar to the generated tile from the DigestPath dataset, we see that the generated tile appears seamless, preserving edge-crossing continuity and morphological characteristics of tissue including epithelial cells, glands and stroma. In addition, although epithelial cells and goblet cells can be clearly distinguished, some moderate deformities in glandular lumen portion are visible in the high resolution image.

\section*{B. Comparison with THeCoT}

We have also compared the visual quality of generated images from SAFRON and the only other existing colorectal tissue image generation method THeCoT by Kovacheva et al. using the same tissue component mask for both methods. This comparison is shown in Fig. 8. The THeCoT parameters used for generating synthetic image are as follows; cellularity of stromal cells = 1, cellularity of epithelial cells = 1, grade of cancer = healthy. We can clearly see that tiles generated from SAFRON are significantly more realistic looking and have better perceptual quality.

\section*{C. Quality Assessment with Frechet Inception Distance}

To assess the proposed framework quantitatively, we compute the Frechet Inception Distance (FID) [25] to evaluate the similarity between the sets of real and generated images. Lower FID corresponds to high perceptual and feature space similarity. For calculating FID, features from the last pooling layer of an Inception V3 network [26] trained on the ”ImageNet” database [27] were used. To get a sense of scale of FID, we use the FID between real images and uniform random noise images of the same size as a baseline. Figure 9 shows FID scores of SAFRON-generated and real images over both DigestPath and CRAG datasets for different tile sizes. Fig. 9 shows low FID between real and SAFRON-generated tiles. This implies that the convolution feature maps computed from SAFRON-generated images are close to the ones obtained for real images. Consequently, it can be concluded that SAFRON-generated images are close to realistic images and can be used in computational pathology applications.

\section*{D. Quality Assessment through Gland Segmentation}

In this experiment, we use a U-net [24] based gland segmentation model to quantify SAFRON image generation quality by comparing gland segmentation outputs for real and corresponding SAFRON-generated images.

For this purpose, we extracted 468 patches of size $512 \times 512$ from CRAG train set to train a U-Net model for gland segmentation. Using the SAFRON framework trained on the CRAG train set, we computed synthetic counterparts of the images in the CRAG test set. We then extracted 471 patches of size $512 \times 512$ from real images as well as SAFRON-generated tiles from the CRAG test set, computed the binary gland segmentation mask on both sets using the trained U-Net model and calculated the Dice index score [28] between the corresponding segmentation outputs as a quality metric. Sample results of segmentation on both real as well as synthetic patches can be seen in the Fig. 10.

We obtained an average Dice index score [28] of 0.93 (with a standard deviation of 0.058) between the computed binary segmentation masks on patches of real and generated tiles. In comparison to an ideal Dice score of 1.0, the score of 0.93 obtained in this experiment shows that there is minimal difference between real and generated tiles for the purposes of gland segmentation.
Fig. 6: Constructed tiles along with their real versions from DP dataset with sizes: (a) 8743 $\times$ 5631 pixels and (b) 6526 $\times$ 5268 pixels
E. Synthetic Images for Training & Performance Evaluation of Machine Learning Models

In this section we describe how SAFRON-generated images can be used for training and performance assessment of a U-Net based gland segmentation model [24] with results at par with evaluation on real images.

For this experiment, we extracted 72 1436 × 1436 pixel tiles from the CRAG test set and divided them into two parts: 40 images for training (U-Net training tiles) and the remaining 32 images for testing (U-Net testing tiles). Using SAFRON trained on the CRAG train set, we generated synthetic counterparts of these 72 tiles. We then extracted two sets of 1930 512 × 512 pixel patches - one from the 40 real U-net training tiles and the other from SAFRON-generated counterparts of these images. These two sets are then used to train the same U-net architecture to yield two different segmentation models - one trained on real images and the other one on SAFRON-generated synthetic images. Each U-net model is trained with 80% of the training images with a learning rate of $10^{-5}$.

We then obtained two test sets of 1639 512 × 512 pixel patches each from the 32 U-net test images and their SAFRON-generated counterparts. These test image patches are not used for training either SAFRON or the U-net segmentation models and are used as unseen test sets for computing Dice scores of binarized segmentation masks outputs with respect to their corresponding ground truth data.

The results of this experiment are shown in table I. It shows that a U-net gland segmentation model trained over real images yields an average Dice score of 0.91 using real test images and 0.90 over SAFRON-generated images. Similarly, a U-net model trained over SAFRON-generated images and tested on real images gives an average Dice score of 0.88 in comparison to 0.97 when synthetic images are used for testing as well. These results clearly show that synthetic images can be used as a replacement of real images in training and performance evaluation of segmentation and possibly other methods in computational pathology.

| Testing \ Training | Real | Synthetic |
|--------------------|------|-----------|
| Real               | 0.91 (0.1) | 0.88 (0.14) |
| Synthetic          | 0.90 (0.1) | 0.97 (0.03) |

TABLE I: Dice index scores obtained after training and testing U-Net on set of patches as shown under the respective column
F. Improving Data Efficiency with Synthetic Data

In this section, we demonstrate how SAFRON-generated images can be used to overcome training data size limitations in training machine learning algorithms by improving their data efficiency.

In this experiment, we track the change in segmentation Dice score of a U-net model over an unseen test set upon augmenting its training dataset with additional real and SAFRON-generated synthetic images. For this purpose, we first divided the 48 large images in the CRAG dataset into three equally sized non-overlapping subsets of 16 images called the Training subset, Test Subset, and Augmentation subset. The test subset is used only for testing the U-net segmentation model and is not used for training either SAFRON or the U-net gland segmentation model. As a baseline, we first trained the U-Net model on 144 patches of size 512 × 512 from the Training subset which yields a Dice score of 0.864 on patches from the test subset which is quite low due to the small amount of training data. We then train SAFRON on 640 square tiles of size 728 × 728 extracted from the same training subset and augment the U-net training dataset with varying numbers of patches from two distinct types of SAFRON-generated synthetic images: one obtained using tissue component masks of images in the augmentation subset and the other based on randomly generated tissue component masks as describe in Section II.C.

The change in segmentation performance of U-net models with different types and number of augmented patches is shown in figure 11. It shows that addition of real images improves the Dice score from 0.864 to 0.93 (with 500 additional patches). However, what is more interesting is the fact that the addition of synthetic images from SAFRON to the U-net training dataset results in a similar improvement even when those images are obtained from randomly generated tissue component masks without using any additional real data either for training SAFRON or the segmentation model.

This experiment clearly demonstrates the usefulness of the proposed scheme in cases where the amount of real annotated training data is small as is frequently the case in machine learning method development in computational pathology.

G. Effect of Skip Connections

In order to understand the effect of skip-connections used in the generator of the network, we evaluated the visual quality of generated images with and without skip-connections.

Fig. 12 shows that the quality of generated images is significantly better when skip connections are used in comparison to when they are not. This shows that skip-connections in the SAFRON framework are crucial for the generation of high quality structures and glandular details.

IV. Conclusions & Future Directions

We presented a novel framework for generation of large histology synthetic image tiles while training on relatively smaller image patches. The constructed tiles from the patch-based framework do not exhibit any boundary artifacts or other deformities between adjacent patches. We showed that the synthetic tissue image tiles generated by SAFRON preserve morphological characteristics in the tissue regions, as demonstrated by the high Dice index for gland segmentation. We demonstrated in this work that the synthetic image tiles constructed from our framework accompanied with a definitive ground-truth generated by a parametric model can be used for pre-training and evaluation of deep learning algorithms in circumstances where labeled data is scarce. SAFRON may be used to extend already existing segmentation datasets for histology image analysis, enabling researchers to improve the performance of automated segmentation approaches for computational pathology. This framework can be generalized for producing a large number of tiles for different types of carcinomas and tissue types. An open future direction for research would be to extend the SAFRON framework to generate complete whole-slide images from known parameters.

REFERENCES

[1] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermen, Q. F. Manson, M. Balkenhol et al., “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[2] T. Qaiser, Y.-W. Tsang, D. Taniyama, N. Sakamoto, K. Nakane, D. Epstein, and N. Rajpoot, “Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features,” *Medical image analysis*, vol. 55, pp. 1–14, 2019.

[3] S. Graham, H. Chen, J. Gamper, Q. Dou, P.-A. Heng, D. Snead, Y.-W. Tsang, and N. Rajpoot, “Mild-net: minimal information loss dilated network for gland instance segmentation in colon histology images,” *Medical image analysis*, vol. 52, pp. 199–211, 2019.
