

Synthesizing Whole Slide Images

Grégory Apou¹, Friedrich Feuerhake², Germain Forestier³, Benoît Naegel¹ and Cédric Wemmert¹

¹University of Strasbourg, ICube, France

²Hannover Medical School, Institute for Pathology, Germany

³University of Haute-Alsace, MIPS, France

Abstract—The increasing availability of digital whole slide images opens new perspectives for computer-assisted image analysis complementing modern histopathology, assuming we can implement reliable and efficient image analysis algorithms to extract the biologically relevant information. Both validation and supervised learning techniques typically rely on ground truths manually made by human experts. However, this task is difficult, subjective and usually not exhaustive. This is a well-known issue in the field of biomedical imaging, and a common solution is the use of artificial “phantoms”. Following this trend, we study the feasibility of synthesizing artificial histological images to create perfect ground truths. In this paper, we show that it is possible to generate a synthetic whole slide image with reasonable computing resources, and we propose a way to evaluate its quality.

I. INTRODUCTION

The multidisciplinary field of digital pathology is a very active research area with major industrial applications in healthcare. According to a recent study [1], the digital pathology market was estimated at \$250.2 million in 2013 with an expected growth of 11.8% by 2018. Despite its current success in telepathology and education [2], its greatest potential has yet to be realized: automated objective medical slide assessment using image processing techniques. The main challenges to tackle are efficient image analysis and data mining techniques, mainly limited by the large size of the images (several billion pixels) and the important variability of both the biological objects and their observation methods [3], [4], [5].

Nowadays, supervised learning is a powerful and convenient way to translate expert knowledge into working software, most notably using Deep Learning techniques such as Convolutional Neural Networks [6]. Unfortunately, in order to provide reliable results despite the biological variability, such supervised machine learning techniques requires large manually validated ground truths that are tedious to produce.

To address this issue, we propose to generate photorealistic Whole Slide Images (WSIs) so that the class of each pixel and the type and location of each cell can be known. Although it does not fully replace the need for expert validation on real data, we strongly believe that it will help to improve the efficiency of the development and testing of automated image analysis tools.

In this paper, we show that it is possible to generate a synthetic WSI with reasonable computing resources, and we propose a way to evaluate its quality (e.g. to what extent the simulated image looks real). A critical analysis of our method is presented along with an exploration of additional techniques

to outline the future steps to further improve the simulation result.

II. MATERIALS

Since the early days of medical imaging, artificial references called “phantoms” have been used to circumvent the numerous difficulties induced by the use of biological material [7]. Modern computers can be used to run virtual simulations of imaging technologies, and the result can be used to validate image analysis algorithm, as it has been done for brain MRI [8], mammography [9] or echocardiography [10].

Most of these tools produce grayscale images, but histological images are stained (colored) to reveal biological objects. When working with them, it can be enough in some cases to make a simple 3D model with only a few cells [11], but our original data consist of diversely stained breast biopsies, with an emphasis on glandular, cancerous and immune tissues and cells. Besides the biological variability, one of the major features of real WSIs is the overall heterogeneity of the stainings and the various artifacts caused by physical manipulation and digitization.

A ground truth in our case would be a biomedically meaningful labeling of an image. An artificial one needs to include all the elements that can impact image analysis, and this implies synthesizing a photorealistic biologically correct high-resolution color image, which is a challenging task. To avoid being overwhelmed by this complexity, we propose to start with a simplified goal: generate a gigapixel-sized image of a normal breast tissue stained with the hematoxylin and eosin standard staining (H&E), and an equivalent image with an immunostaining revealing a type of cells labeled CD8. At the minimum, the image should contain the following elements: background, stroma, adipocytes (fat) and ductal-lobular objects (Figure 1). At first glance, the background is a uniform shade of white. Stroma is composed of spindle-shaped cells and collagen fibers resulting in a longitudinally structured fibrous texture that loses its sharpness during the immunostaining process. Adipocytes are large cells that seem empty. Ductal-lobular objects are the main focus of our study. Having a tree-like 3D structure [12], they induce characteristic 2D patterns: clusters of roundish objects and parts of branching elements.

III. METHODS

A. Image synthesis using region filling

In recent years, powerful methods have been developed to generate visually convincing textures, for instance example-based texture synthesis [13]. The ability to constrain texture

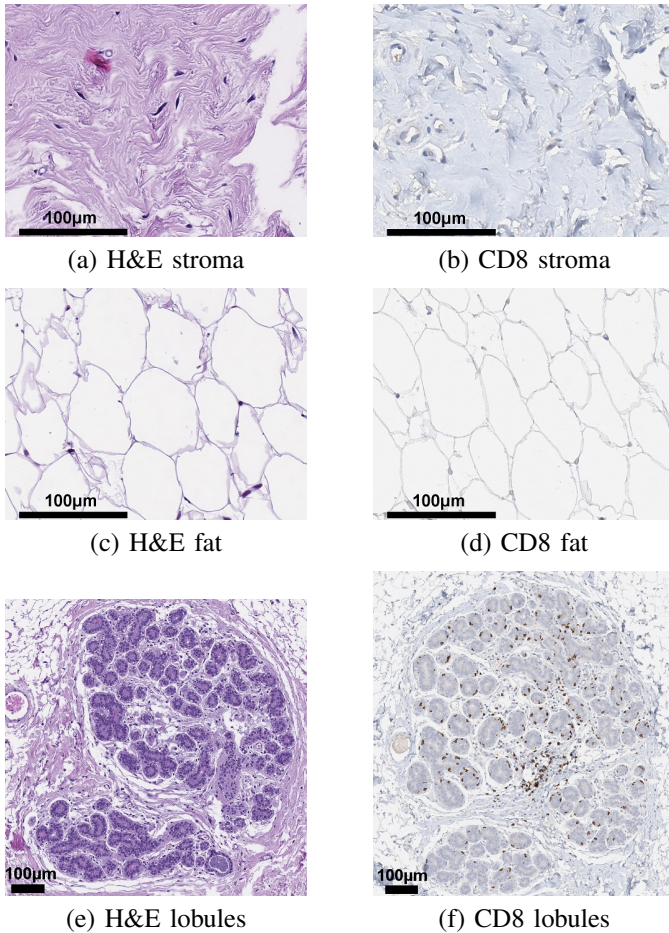


Fig. 1: Main normal breast tissue subtypes in real images. (a), (c), (e): normal staining (H&E). (b), (d), (f): immunostaining (CD8).

synthesis into well defined regions makes it possible to create entire scenes, although such techniques are more often used to modify images [14]. For our purposes, we need to: generate an entire image with billions of pixels; generate different images (stainings) from the same “model”; know the position and type of each visible object; know the origin (object or objects) of each pixel. Additionally, the synthesis of one image should not last too long (a few hours at most); this is a practical constraint because the novelty of our approach requires frequent communications between experts in different fields (medicine and computer science) and this would be facilitated by the possibility of short cycles. The size constraint is particularly problematic because most current software and hardware are not designed to easily handle gigapixel-size images.

The method presented in this paper is designed to fulfill these constraints as simply as possible, so that its weaknesses can be identified and addressed efficiently in future work. It is decomposed in two main steps (Figure 2). In the first step, we start from an empty canvas and we manually define the outlines of regions that are expected to contain a specific label.

Then, in a second step, we fill the drawn areas with textures made of small parts (e.g. cells, ducts, pieces of fiber, etc.) cut from real images (Figure 1).

Before starting the filling process, texture units are manually defined for each class and each staining. A texture unit is an arbitrarily shaped part of an existing image (for example a cell). This is done using the same tool that is used to create the model; the “CD8 stroma” texture is shown in Figure 3.

Algorithm 1 presents how to render a region using a given set of texture units. Associated with the textures, the data structures generated by this algorithm provide a complete and errorless description of the image, from which we can select the relevant information to be exported as a ground truth for a given application.

Algorithm 1 Fill(in: region, in: texture units, out: image)

- 1: compute how many instances of each texture unit are needed to fill region (algorithm 2)
 - 2: initialize the location, collision radius and display orientation of each instance (algorithm 3)
 - 3: **while** instances overlap or time limit not reached **do**
 - 4: re-initialize the location of each instance that lays outside region
 - 5: sort instances by increasing distance to their barycenter
 - 6: move each sorted instance away from each previous instance so that they don’t overlap (algorithm 4)
 - 7: **end while**
 - 8: remove each instance that lays outside region
 - 9: draw each instance into image
-

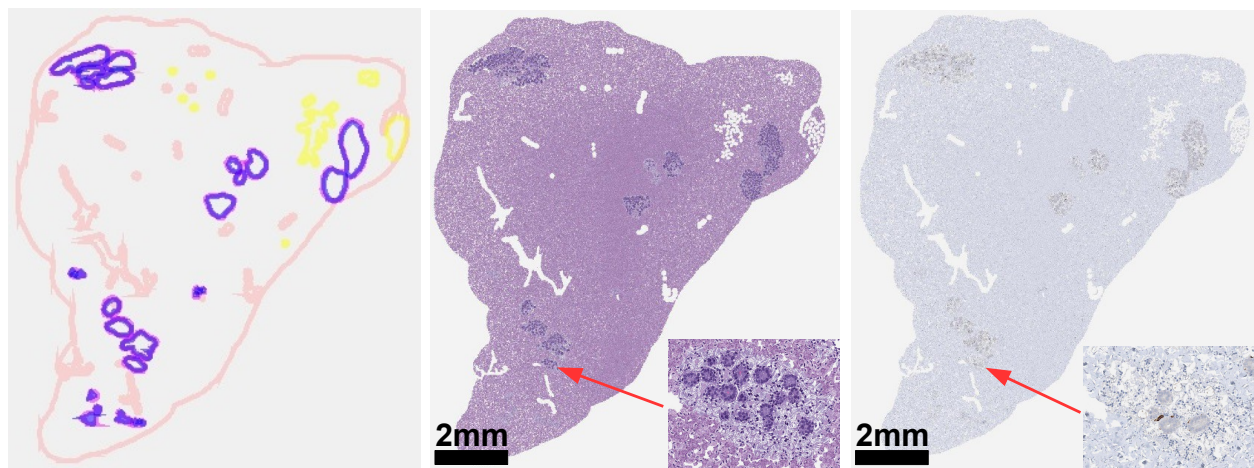
Algorithm 2 EstimateInstances(in: region, in: texture units, out: texture units updated)

- 1: **for** each unit in texture units **do**
 - 2: unit’s proportion \leftarrow random value
 - 3: **end for**
 - 4: normalize texture unit’s proportions so that they sum to 1
 - 5: **for** each unit in texture units **do**
 - 6: unit’s instance count \leftarrow area of region \times unit’s proportion / area of unit
 - 7: **end for**
-

Thanks to its genericity, this method can also be used to simulate other stainings by simply extracting the textures from the appropriate real images.

B. Image synthesis using inpainting or 3D modeling

Given the visual complexity caused by the mixture of periodicity and randomness in histological images, it is natural to inquire about example-based texture synthesis [13] and image completion techniques [14]. Figure 4 illustrates how exemplar-based inpainting [15] can be used to simulate biological tissues.



(a) 2D model of artificial tissue

(b) Synthetic image 1 H&E

(c) Synthetic image 1 CD8

Fig. 2: The 2D model used as input for the simulation (a), and two generated synthetic images for H&E (b) and CB8 (c).

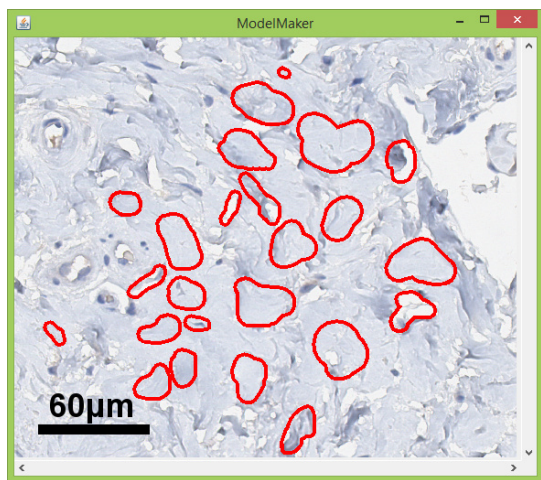


Fig. 3: A simple annotation tool is used to manually define textures; in this image, the pixels of the texture units for the “CD8 stroma” texture are defined by outlining arbitrarily shaped image parts. This step is currently highly subjective but could be easily controlled by establishing guidelines; once we have defined some acceptable metrics, it will be possible to study the impact of the choices made at this step and improve the quality of the generated images.

Algorithm 3 Initialize(in: region, in: instances, out: instances updated)

- 1: **for** each instance in instances **do**
 - 2: instance’s location \leftarrow random location sampled from uniform distribution in disk circumscribing region’s bounding box
 - 3: instance’s radius $\leftarrow \sqrt{\text{unit’s area}/\pi}$
 - 4: instance’s orientation \leftarrow random angle sampled from uniform distribution in $[0, 2\pi]$
 - 5: **end for**
-

Algorithm 4 Collide(in: instances, out: instances updated)

- 1: **for** each i in instances **do**
 - 2: **for** each p in instances before i **do**
 - 3: $v \leftarrow i$ ’s location $- p$ ’s location
 - 4: $c \leftarrow i$ ’s radius $+ p$ ’s radius
 - 5: **if** $\|v\| < c$ **then**
 - 6: i ’s location $\leftarrow p$ ’s location $+ c \times v/\|v\|$
 - 7: **end if**
 - 8: **end for**
 - 9: **end for**
-

However, as our goal is to generate a ground truth, we need to know the location of each synthesized element. For instance, we can grow a 3D model with cell-level details using a branching algorithm (Figure 5 (a) (b)).

To mirror the work done with other imaging technologies, we are working on controlling the appearance of each cell, without having to rely on predefined examples (Figure 5 (c)). Section V discusses why these options have not been used for this paper’s results.

C. Quality assessment

Regardless of the synthesis method used, we need a way to measure the quality of the synthesized images. Unfortunately, there is no universal definition of the subjective concept of photorealism [17]. Moreover, we require some level of biological correctness that only an expert can assess.

We noticed that it can be difficult to distinguish between a real and synthetic image if they are observed at a large scale. This observation is intuitive, as it is generally difficult for a human to differentiate between objects when they are observed at a high level. For example, it is often difficult to differentiate between a real and a fake banknote without looking at them closely. Thus, the reproduction quality of a fake banknote can

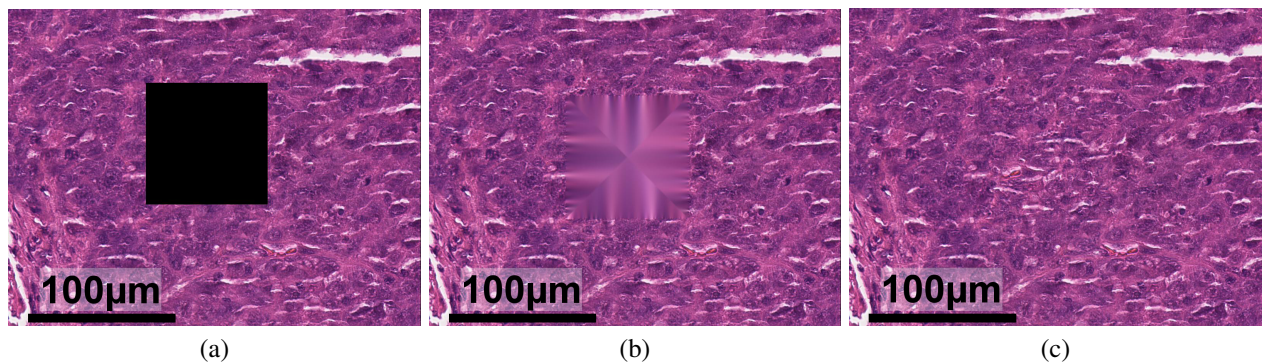


Fig. 4: Texture synthesis by inpainting: (a) Real tissue area (cancerous region, H&E) with “hole”. (b) Hole in (a) badly filled by diffusion-based inpainting [16]. (c) Hole in (a) correctly filled by exemplar-based inpainting [15].

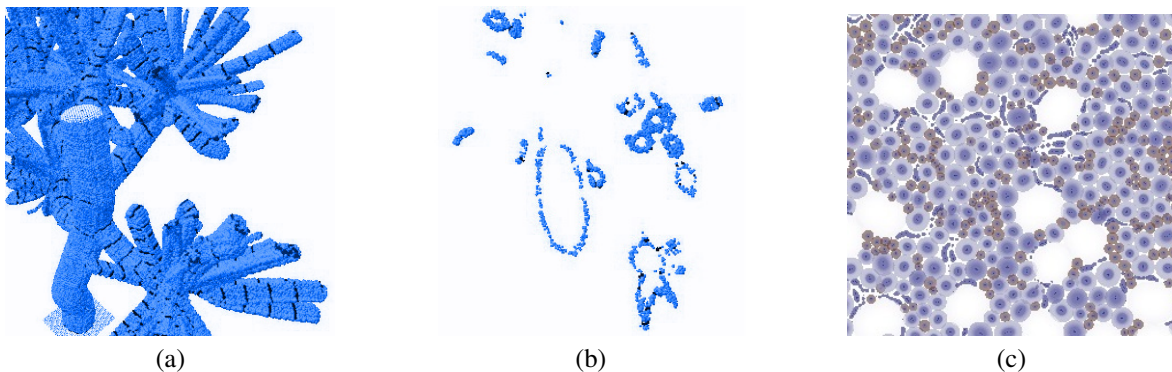


Fig. 5: Procedural structure and texture: (a) Our procedural 3D model of a lobular epithelial layer. (b) Slice of (a) showing 2D lobular patterns. (c) Our procedural cell rendering (random layout, arbitrary colors).

be evaluated using the distance from which you have to be to determine if the note is a fake or not.

Using this intuition, we designed the Algorithm 5 to evaluate the quality of the generated images.

Algorithm 5 Measure realism of synthetic image

- 1: open a synthetic image in a software slide viewer on a standard computer screen
 - 2: decrease the magnification until it “feels realistic”
 - 3: slowly increase the magnification until it “feels wrong”
 - 4: the current magnification is then written down as the highest realistic magnification for the current image
-

IV. RESULTS

Applying the region filling method, we manually defined two sets of outlines inspired from real examples, and also two sets of textures (i.e. one for H&E staining and one for CD8 immunohistochemistry). The four generated pictures were presented to pathologists tasked to look at them by progressively increasing the magnification (Table I).

The highest realistic magnification is the score that evaluates the quality of the image simulation. A high value of this score

Image and staining	Highest realistic magnification	
Image 1 H&E	0.67X	1.09X (0.50)
Image 2 H&E	0.54X	1.09X (0.50)
Average for H&E images	0.61X	1.09X (0.00)
Image 1 CD8	1.22X	0.59X (0.35)
Image 2 CD8	1.06X	0.60X (0.10)
Average for CD8 images	1.14X	0.60X (0.01)

TABLE I: Results before and after including loose stroma; results “after” are given for two pathologists as “average (standard deviation)”. Magnification is a complex concept [18], so we use the following convention: 1X = 10µm/px, 2X = 5µm/px, etc. The highest possible magnification is 40X because the real images from which we extracted our textures were scanned at 40X; the score for a real image should match the scanner’s settings (not tested).

Image (any staining)	Size	Synthesis time
Image 1	30000×40000=1.2 Gpx	20 min
Image 2	70000×70000=4.9 Gpx	90 min

TABLE II: Performance of image synthesis on a desktop computer (Intel quad-core 3.6GHz, RAM 16Gb).

indicates that a pathologist finds the image realistic even by observing it in details.

The times needed to generate the synthetic images are given in Table II.

Using 3 real CD8 images, we performed a small experiment to get a rough idea of the suitability of these synthetic images to evaluate an image analysis method. First, a human expert annotated the lobular structures in each image, giving 3 datasets R1, R2 and R3. Then, we trained a convolutional neural network on R1 (5 iterations for the “CIFAR10 Quick” architecture provided as an example with the Caffe software available from <http://caffe.berkeleyvision.org>, [19]). We evaluated the trained network on R2, R3 and a dataset S built using the synthetic images and their virtual phantoms (as ground truth). The measured accuracies were 0.77, 0.58, 0.56 and 0.68 for R1, R2, R3 and S respectively. These few values seem reasonable and promising.

V. DISCUSSION

The result obtained using our method highlighted that it is possible to synthesize a WSI in an acceptable amount of time.

The visual quality of the synthetic image is satisfactory at low magnification. However, a pathologist is able to see through them as the magnification increases because of the unnatural appearance of the stroma. Moreover, our senior expert noted that the lobular elements were inserted too abruptly into the stroma. This helped us to refine our method and conclude that we should consider two types of stroma: the *normal stroma* and a *loose stroma* that is often present at the periphery of the lobules and allows them to periodically grow and shrink. This assessment helped us to acquire additional biological knowledge that should be used in the synthesis process. The back-and-forth exchanges with the experts is reminiscent of the technique discussed in [20].

The definition of biological photorealism is still an open problem. The way we proposed to measure the photorealism of the images yielded a low score on some synthesized images. The reason given by the experts is that the connective tissue is too dense and homogeneous. However, we have observed that some areas look *credible* at a high magnification, especially the lobules, suggesting that our measure is neither monotonic nor homogeneous. Finally, it is important to remember that we do not aim to build photorealistic images, but to evaluate image processing algorithms. In order to use example-based texture synthesis for that purpose, we would need to carefully annotate the examples, and also make sure that the synthesized structures still represent biological objects. This is currently easier to do by manually drawing the 2D models, but ultimately we hope to automate the whole process.

However, we expect that reproducing the observed 2D patterns correctly requires working with 3D models. We attempted to build a 3D model of a lobular structure made of one layer of cells. Unexpectedly, the current scientific literature does not seem to clearly describe the fine spatial details of human mammary lobules at their different stages of development. Even though this first virtual phantom is

incomplete, it already contains millions of individual cells, and hardware constraints make it unlikely that it can be scaled to contain a useful volume of tissue.

To summarize, this paper presents a method to build synthetic whole slide images that could be used to objectively evaluate image analysis methods. Our future work will involve a combination of different techniques, like exemplar-based inpainting [15] to render the stroma, 3D model slicing to define the ductal-lobular contours, and region filling and procedural cell rendering to finish the image.

VI. ADDITIONAL MATERIAL

Sources, textures and software are available online: <https://github.com/apgrgr/SyntheticWSI>

ACKNOWLEDGMENT

This work was partially funded by the German BMBF as part of the e:Med program.

REFERENCES

- [1] “Digital pathology market by slide scanners (whole slide imaging), analytics (image analysis software), delivery modes (web based/cloud based) & by whole slide image storage - global forecasts & trends to 2018,” <http://www.reportsnreports.com>, 2014.
- [2] F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, “Digital imaging in pathology: Whole-slide imaging and beyond,” *Annual Review of Pathology: Mechanisms of Disease*, vol. 8, no. 1, pp. 331–359, 2013.
- [3] G. Apou, B. Naegel, G. Forestier, F. Feuerhake, and C. Wemmer, “Fast segmentation for texture-based cartography of whole slide images,” in *International Conference on Computer Vision Theory and Applications*, 2014, pp. 309–319.
- [4] J. W. Xu, T. D. Pham, and X. Zhou, “A double thresholding method for cancer stem cell detection,” in *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*, 2011, pp. 695–699.
- [5] P. Fabris, W. Vanzella, and F. A. Pellegrino, “Evaluation of features for automatic detection of cell nuclei in fluorescence microscopy images,” in *IEEE International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2013, pp. 683–688.
- [6] A. Cruz-Roa, A. Basavanahally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi, “Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2014, pp. 904 103–904 103.
- [7] L. A. DeWerd and M. Kissick, Eds., *The Phantoms of Medical and Health Physics*, ser. Biological and Medical Physics, Biomedical Engineering. Springer New York, 2014.
- [8] R.-S. Kwan, A. C. Evans, and G. B. Pike, “MRI simulation-based evaluation of image-processing and classification methods,” *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, 1999.
- [9] K. G. Baum, “Multimodal breast imaging: Registration, visualization, and image synthesis,” Ph.D. dissertation, Rochester Institute of Technology, 2008.
- [10] M. Alessandrini, H. Liebgott, D. Friboulet, and O. Bernard, “Simulation of realistic echocardiographic sequences for ground-truth validation of motion estimation,” in *IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 2329–2332.
- [11] B. Karacali, A. P. Vamvakidou, and A. Tözeren, “Automated recognition of cell phenotypes in histology images based on membrane- and nuclei-targeting biomarkers,” *BMC Medical Imaging*, vol. 7, no. 1, p. 7, 2007.
- [12] J. J. Going and D. F. Moffat, “Escaping from flatland: clinical and biological aspects of human mammary duct anatomy in three dimensions,” *The Journal of Pathology*, vol. 203, no. 1, pp. 538–544, 2004.
- [13] L. Wei, S. Lefebvre, V. Kwatra, and G. Turk, “State of the art in example-based texture synthesis,” in *Eurographics 2009, State of the Art Report, EG-STAR*, 2009, pp. 93–117.
- [14] C. Tang, X. Hu, L. Chen, G. Zhai, and X. Yang, “Sample-based image completion using structure synthesis,” in *IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 1336–1340.

- [15] Anupam, P. Goyal, and S. Diwakar, "Fast and enhanced algorithm for exemplar based image inpainting," in *2010 Fourth Pacific-Rim Symposium on Image and Video Technology (PSIVT)*. IEEE, Nov 2010, pp. 325–330.
- [16] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of graphics tools*, vol. 9, no. 1, pp. 23–34, 2004.
- [17] S. X. D. Højland, C. G. H. Schnatterbeck, K. Jákup, L. L. Østergaard, A. Roman, S. G. Nadarajah, M. W. Hansen, and K. Martin, "Why a single measure of photorealism is unrealistic," in *Proceedings of the Spring Conference on Computer Graphics*, D. Gutierrez and R. Durikovic, Eds. Comenius University, Bratislava, 2014.
- [18] T. L. Sellaro, R. Filkins, C. Hoffman, J. L. Fine, J. Ho, A. V. Parwani, L. Pantanowitz, and M. Montalto, "Relationship between magnification and resolution in digital pathology systems," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 21, 2013.
- [19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.
- [20] R. Bhagavatula, M. T. McCann, M. Fickus, C. A. Castro, J. A. Ozolek, and J. Kovacevic, "A vocabulary for the identification and delineation of teratoma tissue components in hematoxylin and eosin-stained samples," *Journal of pathology informatics*, vol. 5, p. 19, 2014.