Digitally Stained Confocal Microscopy through Deep Learning

Marc Combalia[1] MCOMBALIA@CLINIC.CAT
Javiera Pérez-Anker[1] PEREZ12@CLINIC.CAT
Adriana García-Herrera[2] APGARCIA@CLINIC.CAT
Llúcia Alos[2] LALOS@CLINIC.CAT
Verónica Vilaplana[3] VERONICA.VILAPLANA@UPC.EDU
Ferran Marqués[3] FERRAN.MARQUES@UPC.EDU
Susana Puig[1] SPUIG@CLINIC.CAT
Josep Malvehy[*1] JMALVEHY@CLINIC.CAT

[1] *Dermatology Department, Melanoma Unit, Hospital Clínic de Barcelona, IDIBAPS, Universitat de Barcelona, Barcelona, Spain.*
[2] *Pathology Department, Melanoma Unit, Hospital Clínic de Barcelona, IDIBAPS, Universitat de Barcelona, Barcelona, Spain.*
[3] *Signal Theory and Communications Department, Universitat Politècnica de Catalunya. BarcelonaTech, Spain.*

Editors: Under Review for MIDL 2019

Abstract

Specialists have used confocal microscopy in the ex-vivo modality to identify tumors with an overall sensitivity of 96.6% and specificity of 89.2%. However, this technology hasn't established yet in the standard clinical practice because most pathologists lack the knowledge to interpret its output. In this paper we propose a combination of deep learning and computer vision techniques to digitally stain confocal microscopy images into H&E-like slides, enabling pathologists to interpret these images without specific training. We use a fully convolutional neural network with a multiplicative residual connection to denoise the confocal microscopy images, and then stain them using a Cycle Consistency Generative Adversarial Network.

Keywords: Deep learning, Neural Networks, Digital Staining, Confocal Microscopy, Speckle Noise, CycleGAN

1. Introduction

Histopathology with hematoxylin and eosin (H&E) staining is widely used as a diagnostic tool for a large variety of tissue lesions. However, it requires skilled technicians to process and stain the tissue samples, and it is very costly and time-consuming, requiring from hours to days before a pathologist can analyze the samples. These long delays often impede rapid evaluation of lesions during a surgical operation.

Confocal microscopy (CM) is a novel technique for tissue examination where a laser is focused on a microscopic target and the scattering of the light through its various structures

* Corresponding Author

is captured to form a two-dimensional grayscale image (Calzavara-Pinton et al., 2008). These microscopes can operate in two different modes (reflectance (RCM) and fluorescence (FCM)) which highlight different microscopic structures in the tissue. The combination of the two modes can improve the diagnostic accuracy of the pathologist in the ex-vivo evaluation of tumour margins (Gareau, 2009). In the last years, this new technology has enabled the rapid evaluation of tissue samples directly in the surgery room significantly reducing the time of complex surgical operations in skin cancer.

CMs can obtain images with an optical resolution comparable to pathology, but their output largely differs from the standard H&E slides that pathologist use to evaluate in their clinical practise. Some researchers have focused on creating digitally stained (H&E)-like images from the output of the CMs to facilitate their interpretation by untrained pathologists and surgeons. (Gareau, 2009) has proposed a digital staining technique which linearly combines the FCM and RCM images to form an RGB output slide which resembles H&E stained pathology giving a blue color to FCM and pink to purple color to RCM. This is, in fact, the algorithm used in the last generation of the Vivascope 2500 clinical ex vivo CM device (Vivascope, 2018). This simple staining technique is good at enhancing cellular details allowing the mitosis visualization, but its colors and structures greatly vary from the ones found in the original H&E slides.

In this work, we propose a deep learning technique to combine the two modes of the CM into a (H&E)-like image. First, a fully convolutional neural network is used to remove the speckle noise present in the RCM images, and then a Cycle Consistency Generative Adversarial Network (CycleGAN) (Zhu et al., 2017) is used to combine the FCM and RCM modes into a digitally stained (H&E) slide.

2. Materials and Methods

In this section, we describe the architecture of the Despeckling Neural Network used in the RCM image of the CM and the Generative Adversarial Network used to create the (H&E)-like digitally stained image. Figure 1 shows the complete pipeline for CM image staining.

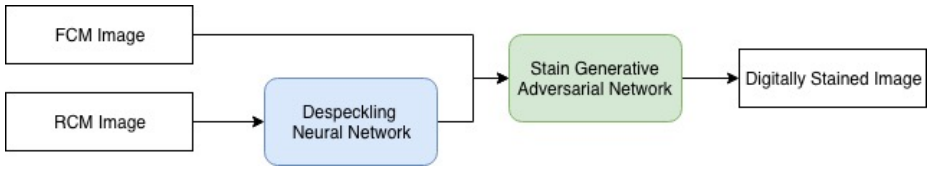


Figure 1: Diagram of the proposed architecture to transform the output of the CM to digitally stained (H&E)-like slides.

2.1. Reflectance Image Despeckling

RCM images are corrupted with a kind of multiplicative noise known as speckle (Sarode and Deshmukh, 2011). Speckle noise is due to a combination of constructive and destructive fluctuations at the input of the CM sensor which interfere with the nominal tissue structure

reflectance. The presence of speckle noise limits the application of further post-processing and computer vision techniques and makes diagnosing less reliable for physicians. Hence, before digitally staining the CM images, their noise must be reduced. Figure 2 shows some RCM images extracted from the CM dataset presented in section 2.3.1.

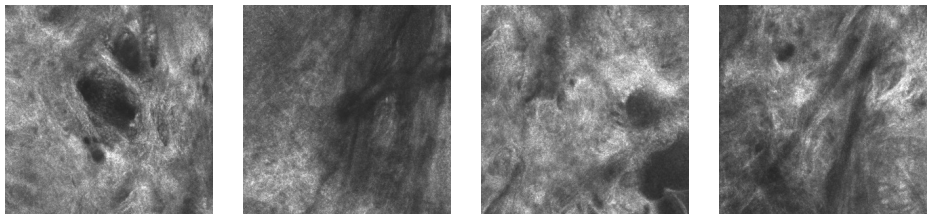


Figure 2: Reflectance images with speckle noise at the output of the CM.

In this work, similarly to (Wang et al., 2017), we use a fully convolutional neural network together with a residual connection to reduce the intensity of the noise present in the RCM images. The observed image at the output of the CM is related to real tissue reflectance by the following equation:

$$Y = X * (1 + F)$$

Where $Y \in \mathcal{R}^{WxH}$ is the observed RCM image, $X \in \mathcal{R}^{WxH}$ is the noise-free reflectance of the tissue, and $F \in \mathcal{R}^{WxH}$ is the speckle noise random variable. We include the aforementioned formulation inside the architecture of the neural network so that it is trained to estimate the inverse of the speckle noise $1/\hat{F}$ at the last convolutional layer.

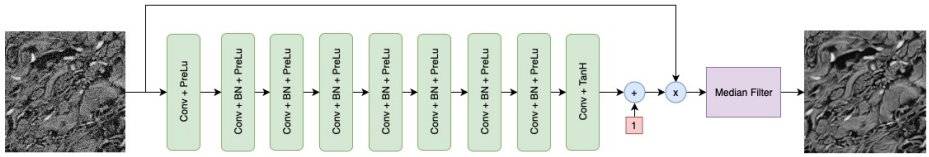


Figure 3: Architecture of the despeckling neural network used in our experiments.

The fully convolutional neural network is composed of 7 convolutional layers (PreLu activations) with 64 filters each and 1 final convolutional layer (TanH activation) with a single filter. Batch normalization is added to the intermediate layers of the neural network. We use a multiplicative residual connection between the last convolutional layer and the input image to incorporate the speckle noise formulation in the the neural network. The architecture is trained to minimize the squared error between the clean images and its output. After training, some noise may still be present in some isolated pixels (figure 6). Authors in (Wang et al., 2017) add total variation loss to the training process to remove these spurious pixel activations. Instead, we filter out the remaining noise using a 3x3 median filter. We train the neural network on a dataset of skin histology images which have been artificially contaminated, which have a similar appearance to noisy RCM images.

2.2. Confocal Microscopy Staining

Due to the impossibility to obtain paired data between the CM domain (A) and the stained H&E histology domain (B) (tissue blocks scanned with the CM need to undergo slicing before staining with H&E), we use Cycle Consistency Generative Adversarial Networks (CycleGAN) (Zhu et al., 2017) to transfer the H&E stain appearance to CM images. The CycleGAN architecture consists of two generator and discriminator pairs. The first pair tries to map images from domain A to domain B, while the second pair undergoes the contrary operation. The generators' task is to create images that the discriminators can't distinguish from real samples. We use a ResNet (He et al., 2016) architecture in the generators and a PatchNet (Isola et al., 2016) in the discriminators. Figure 4 shows all the components and loss functions involved in the translation from A to B.

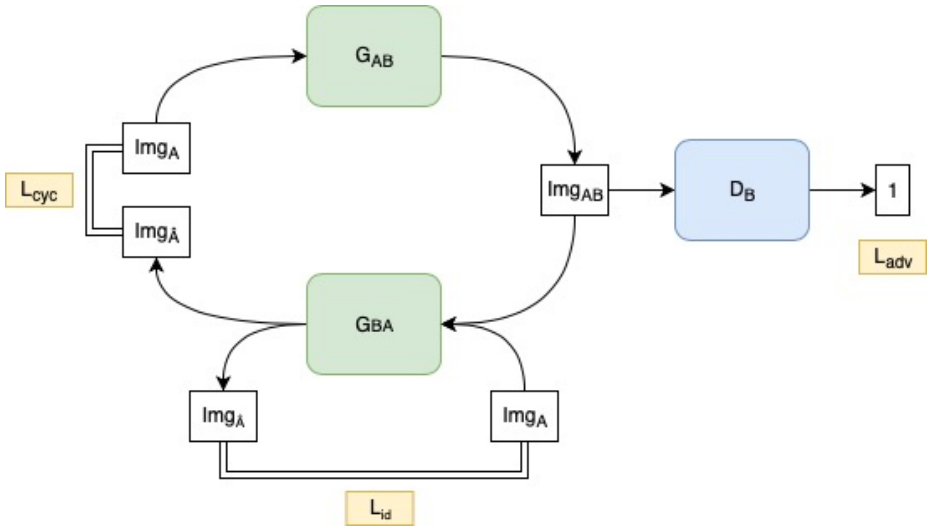


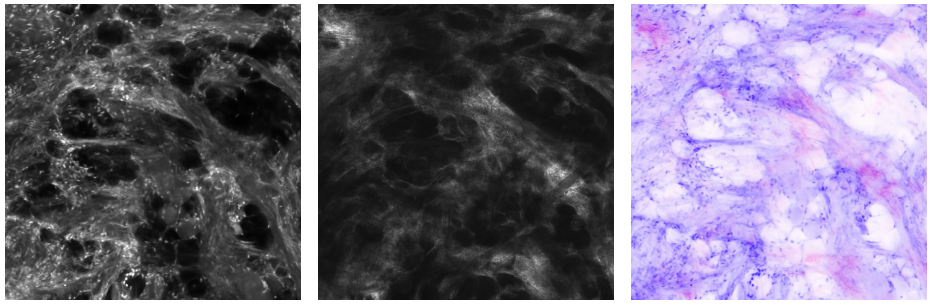
Figure 4: Components and loss functions involved in the domain translation from A to B. The same process is carried out on the contrary direction when translating from B to A.

It is known that CycleGANs are sensitive to their initialization, so to pose an easier task, we use the digital staining method proposed in (Gareau, 2009) as source images for domain A. Figure 5 shows this transformation.

2.3. Data

2.3.1. Confocal Microscopy

Our CM dataset consists of 11 microscopy skin slides obtained with the Vivascope 2500 4th Generation CM (Vivascope, 2018), which captures the tissues at a resolution of 0.75 $\mu m/px$. Its output consists of two large grayscale images (more than 10000 x 10000 pixels), one for each capture mode (RCM and FCM). Both modes are normalized to cover a range from 0 to 1, and OTSU thresholding is used to determine the regions of the slides which



(a) (b) (c)

Figure 5: (a) FCM image, (b) RCM image, (c) Digital staining as proposed by (Gareau, 2009)

contain tissue. Non-overlapping patches of size 1024 x 1024 pixels are extracted summing a total of 949 1024x1024 images for each mode (80 % for the training split and 20 % for the validation split).

2.3.2. H&E Histology

Our H&E Histology dataset consists of 29 skin tissue samples obtained with a Ventana Whole Slide Image scanner captured with a resolution of 0.47 $\mu m/px$. OTSU thresholding is used to determine the tissue containing regions in each whole slide image and multiple patches of size 1630 x 1630 pixels are extracted and then resized to 1024x1024 pixels. The final resolution of each patch is the same as the CM resolution (0.75 $\mu m/px$). The processed dataset consists of a total of 8789 images (80 % for the training split and 20 % for the validation split).

3. Experiments and Results

In this section, we describe the results obtained for the Despeckling Neural Network used in the RCM of the CM and the Generative Adversarial Network used to create the (H&E)-like digitally stained image.

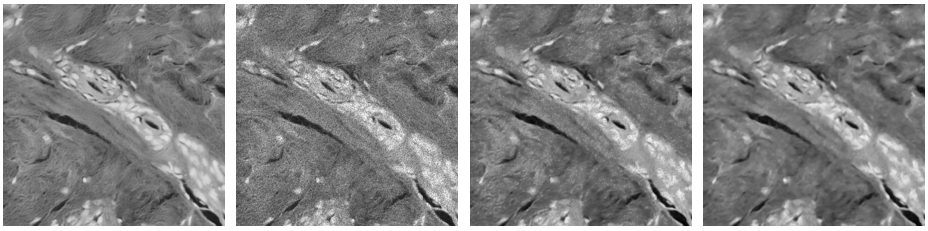
3.1. Reflectance Despeckling

We trained the Desplecking Neural network on 7031 histology images which had been previously transformed to the grayscale color space. We inverted the images to make them closer in appearance to the ones found in the RCM of the CM, where high-intensity pixels correspond to the tissue containing regions. We augmented the training dataset through the use of random flips and random crops of size of 256 x 256 pixels. We updated the weights of the neural network using Adam optimization with a learning rate of $5e-4$. Finally, we evaluated the results on 1748 histology images contaminated with artificial speckle noise and 949 RCM images extracted from the CM. In Table 1 we present the quantitative results obtained on the artificial dataset, and figures 6 and 7 show some images before and after

going through the despeckling neural network for the artificial histology dataset and RCM dataset respectively.

Table 1: PSNR and SSIM before and after applying the proposed Despeckling Neural Network.

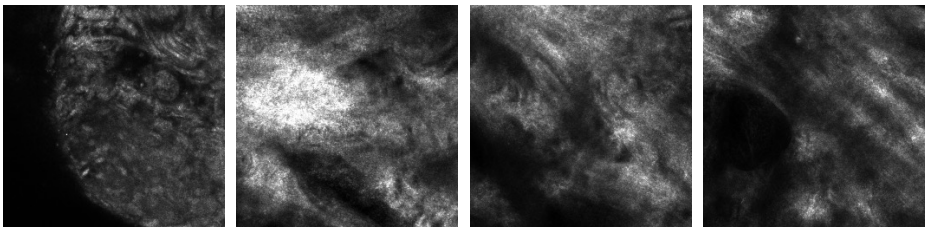
Error Measure	**Noisy**	**Despeckled**
PSNR (dB)	16.19	23.97
SSIM	0.438	0.727



(a) (b) (c) (d)

Figure 6: (a) Original clean image from the artificial despeckling dataset, (b) Original image with artificial speckle noise, (c) Despeckled image at the output of the neural network, before the 3x3 median filter, (d) Despeckled image after the 3x3 median filter.

Original images



Despeckled images

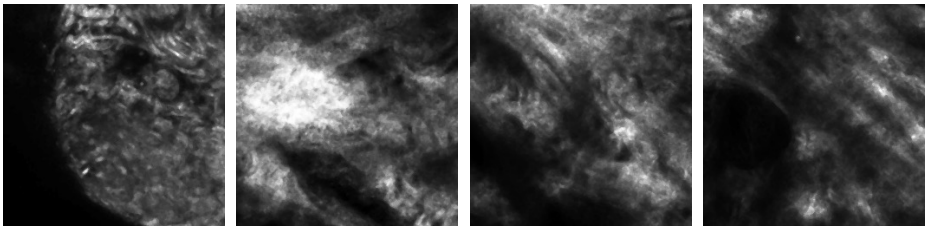


Figure 7: RCM images before and after the proposed despeckling neural network.

3.2. Confocal Image Staining

We trained the Staining CycleGAN on 759 CM images and 282 histology images extracted from a single slide (the one with the best proportion of hematoxylin and eosin stains). The CycleGAN was trained with Adam optimization and a learning rate of $2e-4$. We trained the neural network first on patches of size 256 x 256 pixels. After 50 epochs, we augmented the patch size to 512 x 512 so that the architecture could learn features seen at a higher scale and then trained it for another 50 epochs with learning rate decay. Figure 8 shows some results of a CycleGAN trained on images which have been previously denoised with the method described in section 2.1, as well as some images from domain B. Figure 9 shows some results of a CycleGAN trained with noisy RCM images.

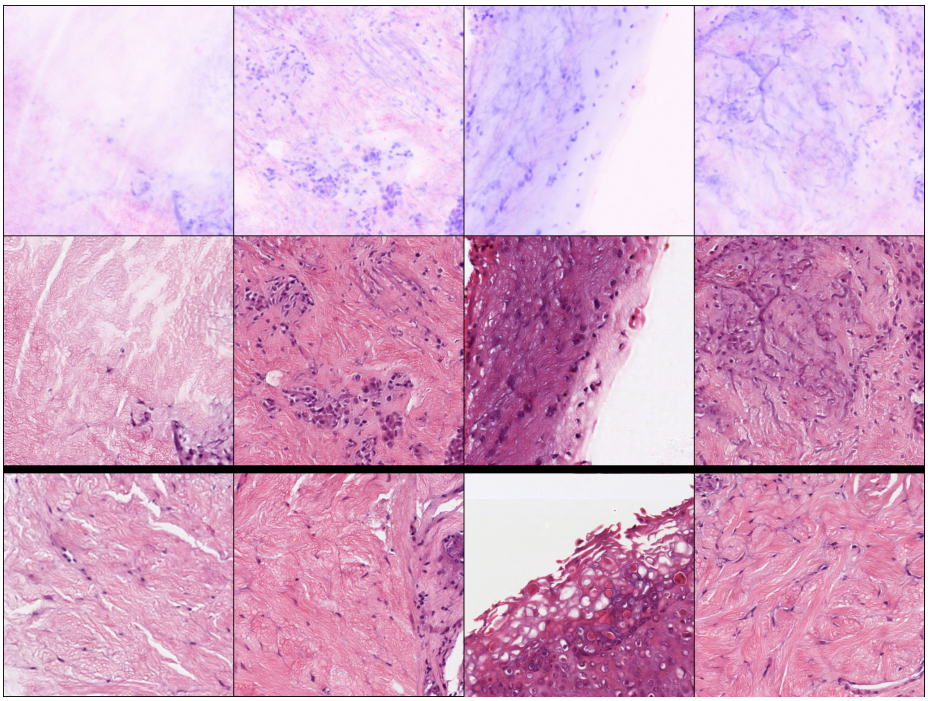


Figure 8: Results of the proposed architecture (Despeckling neural network and CycleGAN). The top row represents the input images of the CycleGAN, which have been digitally stained with the method proposed by (Gareau, 2009). Middle row is the output of the staining CycleGAN. The images in the bottom row are real H&E stained histology images extracted from the training dataset. All images are 512x512 pixels.

4. Discussion

We argue that the combination of the proposed despeckling neural network with the CycleGAN architecture for stain transfer is capable of producing realistic (H&E)-like images.

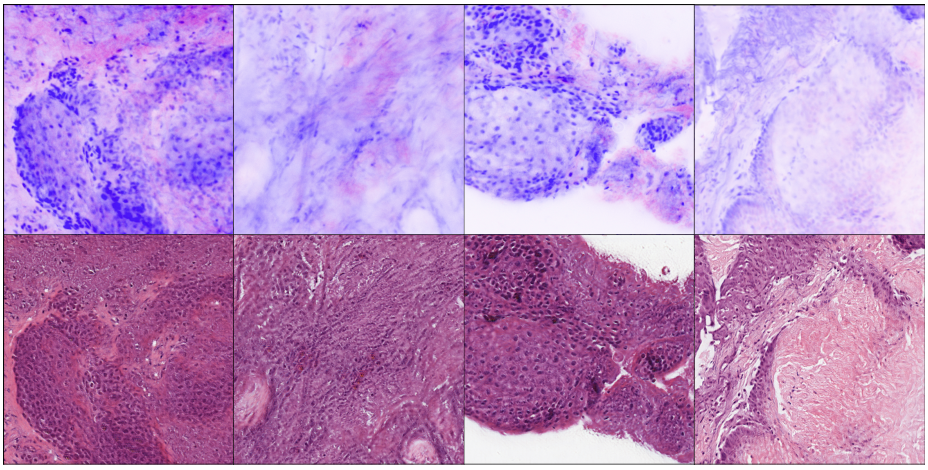


Figure 9: Results from the CycleGAN trained with RCM noisy images. Top row represents the input images of the CycleGAN, which have been digitally stained with the method proposed by (Gareau, 2009). Bottom row is the output of the staining CycleGAN. All images are 512x512 pixels.

Output images from the proposed algorithm were evaluated by two expert pathologists in our department (LL.A/A.G) and they confirmed that the images were similar to those in routine.

The despeckling neural network was able to successfully remove the noise from the RCM images at the output of the CM. From the results on figure 9 we conclude that the Despeckling Neural Network is crucial to obtain realistic images at the output of the CycleGAN. The architecture trained with noisy RCM images had a harder time learning to map the confocal output to the (H&E)-like appearance and produced non-desirable artifacts, as well as eliminated some nuclei present in the CM images. However, we argue that the Despeckling Neural Network could benefit from including an adversarial loss in the optimization process to produce sharper results.

5. Conclusions and Future Work

We have proposed an architecture which successfully addresses the problems involved in CM image staining. On the one hand, we reduce the noise present in the RCM images through the use of a denoising convolutional neural network with a multiplicative residual connection. Then, the denoised RCM image and FCM image are combined in a generative adversarial network to produce a realistic (H&E)-like output image. The methods described in this paper will undergo clinical validation in the near future.

References

Piergiacomo Calzavara-Pinton, Caterina Longo, Marina Venturini, Raffaella Sala, and Giovanni Pellacani. Reflectance confocal microscopy for in vivo skin imaging. *Photochemistry and photobiology*, 84(6):1421–1430, 2008.

Daniel S Gareau. Feasibility of digitally stained multimodal confocal mosaics to simulate histopathology. *Journal of biomedical optics*, 14(3):034050, 2009.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. URL http://arxiv.org/abs/1611.07004.

M.V. Sarode and P.R. Deshmukh. Reduction of speckle noise and image enhancement of images using filtering technique. *International Journal of Advancements in Technology*, 2:30–38, 01 2011.

Vivascope. Vivascope. http://www.vivascope.de/home.html, 2018. Accessed: 2018-12-13.

Puyang Wang, He Zhang, and Vishal M Patel. Sar image despeckling using a convolutional neural network. *IEEE Signal Processing Letters*, 24(12):1763–1767, 2017.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017. URL http://arxiv.org/abs/1703.10593.