Objective: take unstained auto-flovescence images as input and generate stained bright-field images.

General Architecture: GAN

Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue

Yair Rivenson[1,2,3†], Hongda Wang[1,2,3†], Zhensong Wei[1], Yibo Zhang[1,2,3], Harun Günaydın[1], Aydogan Ozcan[1,2,3,4,*]

[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA, 90095, USA

[2]Bioengineering Department, University of California, Los Angeles, CA, 90095, USA

[3]California NanoSystems Institute, University of California, Los Angeles, CA, 90095, USA

[4]Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA.

†Equal contributing authors.

*Email: ozcan@ucla.edu

ABSTRACT

Histological analysis of tissue samples is one of the most widely used methods for disease diagnosis. After taking a sample from a patient, it goes through a lengthy and laborious preparation, which stains the tissue to visualize different histological features under a microscope. Here, we demonstrate a label-free approach to create a virtually-stained microscopic image using a single wide-field auto-fluorescence image of an unlabeled tissue sample, bypassing the standard histochemical staining process, saving time and cost. This method is based on deep learning, and uses a convolutional neural network trained using a generative adversarial network model to transform an auto-fluorescence image of an unlabeled tissue section into an image that is equivalent to the bright-field image of the stained-version of the same sample. We validated this method by successfully creating virtually-stained microscopic images of human tissue samples, including sections of salivary gland, thyroid, kidney, liver and lung tissue, also covering three different stains. This label-free virtual-staining method eliminates cumbersome and costly histochemical staining procedures, and would significantly simplify tissue preparation in pathology and histology fields.

Microscopic imaging of tissue samples is a fundamental tool used for the diagnosis of various diseases and forms the workhorse of pathology and biological sciences. The clinically-established gold standard image of a tissue section is the result of a laborious process, which includes the tissue specimen being formalin-fixed paraffin-embedded (FFPE), sectioned to thin slices (typically ~2-10µm), labeled/stained and mounted on a glass slide, which is then followed by its microscopic imaging using e.g., a bright-field microscope. All these steps use multiple reagents and introduce irreversible effects on the tissue. There have been recent efforts to change this workflow using different imaging modalities. One line of work imaged fresh, non-paraffin-embedded tissue samples using non-linear microscopy methods based on e.g., two-photon fluorescence, second harmonic generation[1], third-harmonic generation[2] as well as Raman scattering[3,4]. Another study used a controllable super-continuum source[5] to acquire multi-modal images for chemical analysis of fresh tissue samples. These methods require using ultra-fast lasers or super-continuum sources, which might not be readily available in most settings and require relatively long scanning times due to weaker optical signals. In addition to these, other microscopy methods for imaging non-sectioned tissue samples have also emerged by using UV-excitation on stained samples[6,7], or by taking advantage of the auto-fluorescence emission of biological tissue at short wavelengths[8]. In fact, auto-fluorescence signal creates some unique opportunities for imaging tissue samples by making use of the fluorescent light emitted from endogenous fluorophores. It has been demonstrated that such endogenous fluorescence signatures carry useful information that can be mapped to functional and structural properties of biological specimen and therefore have been used extensively for diagnostics and research purposes[8–10]. One of the main focus areas of these efforts has been the spectroscopic investigation of the relationship between different biological molecules and their structural properties under different conditions. Some of these well characterized biological constitutes include vitamins (e.g., vitamin A, riboflavin, thiamin), collagen, coenzymes, fatty acids, among others[9].

While some of the above discussed techniques have unique capabilities to discriminate e.g., cell types and sub-cellular components in tissue samples using various contrast mechanisms, pathologists as well as tumor classification software[11] are in general trained for examining histochemically stained tissue samples to make diagnostic decisions. Partially motivated by this, some of the above mentioned techniques were also augmented to create pseudo-Hematoxylin and Eosin (H&E) images[1,12], which were based on a linear approximation that relates the fluorescence intensity of an image to the dye concentration per tissue volume, using empirically determined constants that represent the mean spectral response of various dyes embedded in the tissue. These methods also used exogenous staining to enhance the fluorescence signal contrast in order to create virtual H&E images of tissue samples.

In this work, we demonstrate deep learning-based virtual histology staining using auto-fluorescence of unstained tissue, imaged with a wide-field fluorescence microscope through a standard near-UV excitation/emission filter set (see the Methods section). The virtual staining is performed on a single auto-fluorescence image of the sample by using a deep Convolutional Neural Network (CNN), which is trained using the concept of Generative Adversarial Networks (GAN)[13] to match the bright-field microscopic images of tissue samples after they are labeled with a certain histology stain (see Fig. 1 and Supplementary Figs. S1-S2). Therefore, using a

CNN, we replace the histochemical staining and bright-field imaging steps with the output of the trained neural net, which is fed with the auto-fluorescence image of the unstained tissue. The network inference is fast, taking e.g., ~0.59 sec using a standard desktop computer for an imaging field-of-view of ~ 0.33 mm × 0.33 mm using e.g., a 40× objective lens.

We demonstrated this deep learning-based virtual histology staining method by imaging label-free human tissue samples including salivary gland, thyroid, kidney, liver and lung, where the network output created equivalent images, very well matching the images of the same samples that were labeled with three different stains, i.e., H&E (salivary gland and thyroid), Jones stain (kidney) and Masson's Trichrome (liver and lung). Since the network's input image is captured by a conventional fluorescence microscope with a standard filter set, this approach has transformative potential to use unstained tissue samples for pathology and histology applications, entirely bypassing the histochemical staining process, saving time and cost. For example, for the histology stains that we learned to virtually stain in this work, each staining procedure of a tissue section on average takes ~45 min (H&E) and 2-3 hours (Masson's Trichrome and Jones stain), with an estimated cost, including labor, of \$2-5[14,15] (H&E) and >\$16-35[15,16] (Masson's Trichrome and Jones stain). Furthermore, some of these histochemical staining processes require time-sensitive steps, demanding the expert to monitor the process under a microscope, which makes the entire process not only lengthy and relatively costly, but also laborious. The presented method bypasses all these staining steps, and also allows the preservation of unlabeled tissue sections for later analysis, such as micro-marking of sub-regions of interest on the unstained tissue specimen that can be used for more advanced immunochemical and molecular analysis to facilitate e.g., customized therapies[17,18]. Also note that, this deep learning-based virtual histology staining framework can be broadly applied to other excitation wavelengths or fluorescence filter sets, as well as to other microscopy modalities (such as non-linear microscopy) that utilize additional endogenous contrast mechanisms. In our experiments, we used sectioned and fixed tissue samples to be able to provide meaningful comparisons to the results of the standard histochemical staining process. However, the presented approach would also work with non-fixed, non-sectioned tissue samples, potentially making it applicable to use in surgery rooms or at the site of a biopsy for rapid diagnosis or telepathology applications. Beyond its clinical applications, this method could broadly benefit histology field and its applications in life science research and education.

RESULTS

Virtual staining of tissue samples

We demonstrated the presented method using different combinations of tissue sections and stains. Following the training of a deep CNN (outlined in the Methods Section) we blindly tested its inference by feeding it with the auto-fluorescence images of label-free tissue sections that did not overlap with the images that were used in the training or validation sets. Figure 2 summarizes our results for a salivary gland tissue section, which was virtually stained to match H&E stained bright-field images of the same sample. These results demonstrate the capability of the presented framework to transform an auto-fluorescence image of a label-free

strong structural similarity between the network output images and the bright-field images of the chemically stained samples.

One should note that the bright-field images of the chemically stained tissue samples in fact do not provide the *true* gold standard for this specific comparison of the network output, because there are uncontrolled variations and structural changes (see e.g., Supplementary Fig. 3) that the tissue undergoes during the histochemical staining process and related dehydration and clearing steps. Another variation that we noticed for some of the images was that the automated microscope scanning software selected different auto-focusing planes for the two imaging modalities. All these variations create some challenges for the absolute quantitative comparison of the two sets of images (i.e., the network output for a label-free tissue vs. the bright-field image of the same tissue after the histological staining process). We further expand this point in the Discussion section.

Transfer learning to other tissue-stain combinations

Using the concept of transfer learning[20], the training procedure for new tissue and/or stain types can converge much faster, while also reaching an improved performance, i.e., a better local minimum in the training cost/loss function (see the Methods section). This means, a pre-learnt CNN model, from a different tissue-stain combination, can be used to initialize the deep network to statistically learn virtual staining of a new combination. Figure 5 demonstrates the favorable attributes of such an approach: a new deep neural network was trained to virtually stain the auto-fluorescence images of unstained *thyroid* tissue sections, and it was initialized using the weights and biases of another network that was previously trained for H&E virtual staining of the *salivary gland*. The evolution of the loss metric as a function of the number of iterations used in the training phase clearly demonstrates that the new thyroid deep network rapidly converges to a lower minimum in comparison to the same network architecture which was trained from scratch, using random initialization. Figure 5 also compares the output images of this thyroid network at different stages of its learning process, which further illustrates the impact of transfer learning to rapidly adapt the presented approach to new tissue/stain combinations. The network output images, after the training phase with e.g., $\geq$ 6,000 iterations, reveal that cell nuclei show irregular contours, nuclear grooves, and chromatin pallor, suggestive of papillary thyroid carcinoma; cells also show mild to moderate amounts of eosinophilic granular cytoplasm and the fibrovascular core at the network output image shows increased inflammatory cells including lymphocytes and plasma cells.

DISCUSSION

We demonstrated the ability to virtually stain label-free tissue sections, using a supervised deep learning technique that uses a single auto-fluorescence image of the sample as input, captured by a standard fluorescence microscope and filter set. This statistical learning-based method has the potential to restructure the clinical workflow in histopathology and can benefit from various imaging modalities such as fluorescence microscopy, non-linear microscopy, holographic microscopy and optical coherence tomography[21], among others, to potentially provide a digital

Following this FOV matching procedure, the auto-fluorescence and bright-field microscope images are coarsely matched. However, they are still not accurately registered at the individual pixel-level, due to the slight mismatch in the sample placement at the two different microscopic imaging experiments (auto-fluorescence, followed by bright-field), which randomly causes a slight rotation angle (e.g., ~1-2 degrees) between the input and target images of the same sample.

The second part of our input-target matching process involves a global registration step, which corrects for this slight rotation angle between the auto-fluorescence and bright-field images. This is done by extracting feature vectors (descriptors) and their corresponding locations from the image pairs, and matching the features by using the extracted descriptors[24]. Then, a transformation matrix corresponding to the matched pairs is found using the M-estimator Sample Consensus (MSAC) algorithm[25], which is a variant of the Random Sample Consensus (RANSAC) algorithm[26]. Finally, the angle-corrected image is obtained by applying this transformation matrix to the original bright-field microscope image patch. Following the application of this rotation, the images are further cropped by 100 pixels (50 pixels on each side) to accommodate for undefined pixel values at the image borders, due to the rotation angle correction.

Finally, for the local feature registration we applied an elastic image registration algorithm, which matches the local features of both sets of images (auto-fluorescence vs. bright-field), by hierarchically matching the corresponding blocks, from large to small (see Supplementary Fig. S5). The calculated transformation map from this step is finally applied to each bright-field image patch[27].

At the end of these registration steps, the auto-fluorescence image patches and their corresponding bright-field tissue image patches are accurately matched to each other and can be used as input and label pairs for the deep neural network training phase, allowing the network to *solely* focus on and learn the problem of virtual histological staining.

Deep neural network architecture and training

In this work, we used a GAN[13] architecture to learn the transformation from a label-free unstained auto-fluorescence input image to the corresponding bright-field image of the chemically stained sample. A standard convolutional neural network-based training learns to minimize a loss/cost function between the network's output and the target label. Thus, the choice of this loss function is a critical component of the deep network design. For instance, simply choosing an ℓ_2-norm penalty as a cost function will tend to generate blurry results[28,29], as the network averages a weighted probability of all the plausible results; therefore, additional regularization terms[30,31] are generally needed to guide the network to preserve the desired sharp sample features at the network's output. GANs avoid this problem by learning a criterion that aims to accurately classify if the deep network's output image is real or fake (i.e., correct in its virtual staining or wrong). This makes the output images that are inconsistent with the desired labels not to be tolerated, which makes the loss function to be *adaptive* to the data and the desired task at hand. To achieve this goal, the GAN training procedure involves training of two different networks, as shown in Supplementary Figs. 1-2: (*i*) a *generator* network, which in our case aims to learn the statistical transformation between the unstained auto-fluorescence input images and the corresponding bright-field images of the same samples, after the histological

staining process; and (*ii*) a *discriminator* network that learns how to discriminate between a true bright-field image of a stained tissue section and the generator network's output image. Ultimately, the desired result of this training process is a generator, which transforms an unstained auto-fluorescence input image into an image which will be *indistinguishable* from the stained bright-field image of the same sample. For this task, we defined the loss functions of the generator and discriminator as such:

$$\begin{aligned} \ell_{\text{generator}} &= \text{MSE}\left\{z_{\text{label}}, z_{\text{output}}\right\} + \lambda \times \text{TV}\left\{z_{\text{output}}\right\} + \alpha \times (1 - D(z_{\text{output}}))^2 \\ \ell_{\text{discrimnator}} &= D(z_{\text{output}})^2 + (1 - D(z_{\text{label}}))^2 \end{aligned} \tag{1}$$

where D refers to the discriminator network output, z_{label} denotes the bright-field image of the chemically stained tissue, z_{output} denotes the output of the generator network. The generator loss function balances the pixel-wise mean squared error (MSE) of the generator network output image with respect to its label, the total variation (TV) operator of the output image, and the discriminator network prediction of the output image, using the regularization parameters (λ, α) that are empirically set to different values, which accommodate for ~2% and ~20% of the pixel-wise MSE loss and the combined generator loss ($\ell_{\text{generator}}$), respectively. The TV operator of an image z is defined as:

$$\text{TV}(z) = \sum_p \sum_q \sqrt{\left(z_{p+1,q} - z_{p,q}\right)^2 + \left(z_{p,q+1} - z_{p,q}\right)^2} \tag{2}$$

where p, q are pixel indices. Based on Eq. (1), the discriminator attempts to minimize the output loss, while maximizing the probability of correctly classifying the real label (i.e., the bright-field image of the chemically stained tissue). Ideally, the discriminator network would aim to achieve $D(z_{\text{label}}) = 1$ and $D(z_{\text{output}}) = 0$, but if the generator is successfully trained by the GAN, $D(z_{\text{output}})$ will ideally converge to 0.5.

The generator deep neural network architecture follows the design of U-net[32], and is detailed in Supplementary Fig. S2. An input image is processed by the network in a multi-scale fashion, using down-sampling and up-sampling paths, helping the network to learn the virtual staining task at various different scales. The down-sampling path consists of four individual steps, with each step containing one residual block[33], each of which maps a feature map x_k into feature map x_{k+1}:

$$x_{k+1} = x_k + \text{LReLU}\left[\text{CONV}_{k3}\left\{\text{LReLU}\left[\text{CONV}_{k2}\left\{\text{LReLU}\left[\text{CONV}_{k1}\left\{x_k\right\}\right]\right\}\right]\right\}\right] \tag{3}$$

where CONV{.} is the convolution operator (which includes the bias terms), *k1*, *k2*, and *k3* denote the serial number of the convolution layers, and LReLU[.] is the non-linear activation function (i.e., a Leaky Rectified Linear Unit) that we used throughout the entire network, defined as: