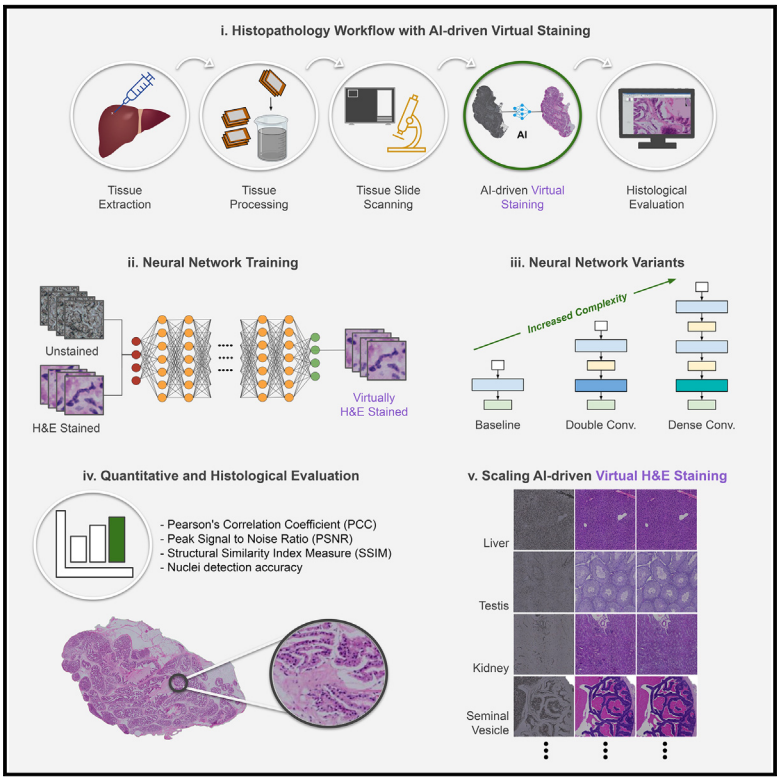
Article

Patterns

The effect of neural network architecture on virtual H&E staining: Systematic assessment of histological feasibility

Graphical abstract



Highlights

- Deep learning-based virtual staining of unstained histological tissue is presented
- We show that increasing the network capacity produces better virtual staining
- Our method uses regular brightfield microscopy and is widely applicable
- Virtual staining has great potential in reducing the need for chemical staining

Authors

Umair Khan, Sonja Koivukoski, Mira Valkonen, Leena Latonen, Pekka Ruusuvuori

Correspondence

pekka.ruusuvuori@utu.fi

In brief

Conventional histopathology uses chemical staining as the gold standard for tissue analysis, but it is a time-intensive, laborious, and irreversible process. This study systematically evaluates the potential for deep neural networks in the virtual staining of tissue images obtained with regular brightfield microscopy. For tissues from multiple organs, we performed quantitative and visual evaluation of the reproduction accuracy of virtual staining vs. H&E-stained ground truth. Using variants of the generative adversarial network model pix2pix, we show that increasing neural network complexity can lead to higher virtual staining quality. Our study suggests that virtual staining could be used to reduce the need for chemical staining in histopathology.

Khan et al., 2023, Patterns *4*, 100725
May 12, 2023
https://doi.org/10.1016/j.patter.2023.100725

CellPress

Patterns

CellPress
OPEN ACCESS

Article

The effect of neural network architecture on virtual H&E staining: Systematic assessment of histological feasibility

Umair Khan,[1] Sonja Koivukoski,[2] Mira Valkonen,[3] Leena Latonen,[2,4] and Pekka Ruusuvuori[1,3,5,6,*]
[1]University of Turku, Institute of Biomedicine, Turku 20014, Finland
[2]University of Eastern Finland, Institute of Biomedicine, Kuopio 70211, Finland
[3]Tampere University, Faculty of Medicine and Health Technology, Tampere 33100, Finland
[4]Foundation for the Finnish Cancer Institute, Helsinki 00290, Finland
[5]FICAN West Cancer Centre, Cancer Research Unit, Turku University Hospital, Turku 20500, Finland
[6]Lead contact
*Correspondence: pekka.ruusuvuori@utu.fi
https://doi.org/10.1016/j.patter.2023.100725

THE BIGGER PICTURE Virtual staining of unstained histological tissue by using deep neural networks has the potential to streamline the sample processing phase in histopathology and to reduce material consumption. The applicability of virtual staining as a replacement for traditional staining, however, is dependent on its accuracy in repeating the staining patterns on macro and cellular levels. Here, we show through quantitative and comprehensive visual evaluation of tissue samples from several organs that increasing the capacity of the neural networks produces better virtual staining with fewer artifacts. Our study suggests that AI-enabled virtual staining of unstained tissue obtained using a widely available basic brightfield microscopy setup can be used to potentially omit the staining process. This technology is scalable and has tremendous potential in improving sustainability, enabling savings in laboratory work, chemicals, and in use of histological tissue specimens.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Conventional histopathology has relied on chemical staining for over a century. The staining process makes tissue sections visible to the human eye through a tedious and labor-intensive procedure that alters the tissue irreversibly, preventing repeated use of the sample. Deep learning-based virtual staining can potentially alleviate these shortcomings. Here, we used standard brightfield microscopy on unstained tissue sections and studied the impact of increased network capacity on the resulting virtually stained H&E images. Using the generative adversarial neural network model pix2pix as a baseline, we observed that replacing simple convolutions with dense convolution units increased the structural similarity score, peak signal-to-noise ratio, and nuclei reproduction accuracy. We also demonstrated highly accurate reproduction of histology, especially with increased network capacity, and demonstrated applicability to several tissues. We show that network architecture optimization can improve the image translation accuracy of virtual H&E staining, highlighting the potential of virtual staining in streamlining histopathological analysis.

INTRODUCTION

Histological stainings are used to colorize tissue specimens, making the almost transparent tissue sections visible. The chemical staining process is currently required for standard pathological observations in disease research and in clinical diagnostics. Different types of dyes manifest different colors in the stained tissue, adding contrast and revealing details such as cellular and sub-cellular morphological information that is otherwise indiscernible in unstained tissue. For instance, hematoxylin and eosin

(H&E), first introduced in 1876 by A. Wissowzky,[1] is one of the most commonly used stain combinations; hematoxylin gives cell nuclei a purplish-blue color, and eosin gives the extracellular matrix and cytoplasm different shades of pink.[2]

While conventional histopathology staining has been the gold standard for tissue analysis for decades, it comes with its fair share of drawbacks. The staining process is time- and chemical-consuming, and resource-intensive. The staining protocols and color manifestation of dyes vary from one laboratory to another.[3] Moreover, the histological process is laborious and includes several manual phases where potential technical variations or artifacts may be introduced. Further, the current chemical protocols allow only one staining to be performed per tissue section. Hence, for multiple stainings, additional tissue sections are required. This adds to the consumption of often limited tissue samples in, e.g., clinical diagnostics.

A possible solution to these problems lies in virtualizing the staining process. Virtual staining refers to employing an algorithmic approach that takes the scanned image of an unstained tissue specimen as input and generates its corresponding stained version digitally. In computer vision, this problem, in general, is known as image-to-image translation and there are several different methods to solve it.[4] In medical imaging, image-to-image translation methods have been used especially in radiology for cross-modality synthesis, for instance, magnetic resonance (MR) to computed tomography (CT),[5,6] MR T1-weighted to T2-weighted sequence, and vice versa.[7] These methods have rather quickly made the transition from research to Food and Drug Administration- and European Medicines Agency-approved commercial products meant for clinical use.[8,9]

In recent years, deep learning has greatly impacted the field of computer vision. Deep learning-based algorithms have achieved superior performance over conventional machine learning methods at tasks such as image classification and segmentation,[10–13] and image-to-image translation is no exception here. Most of the deep learning-based image-to-image translation methods stem from an image synthesis method called generative adversarial network (GAN).[14] A GAN uses two models: a generator and a discriminator, and both models are trained in a game theoretic way. The models compete in a zero-sum game, where the generator tries to generate as realistic synthetic images as possible by learning a mapping function of a latent space to the target domain, whereas the discriminator tries to distinguish synthetic images from the real ones by learning to distinguish features of both types of images. The method attains its learning objective when the realism of generated or synthetic images reaches a point where the discriminator can no longer tell synthetic images apart from the real ones, and the generator cannot further improve the realism of the synthetic images. This phenomenon is known as Nash equilibrium in game theory.[15]

Image-to-image translation methods are mainly divided into supervised and unsupervised categories. Supervised image-to-image translation methods require aligned or registered image data with pixel-to-pixel correspondence for the training phase, whereas unsupervised methods typically use distribution matching loss functions such as cycle consistency loss,[16] instead of relying on pixel-wise correspondence. Unsupervised methods such as CycleGAN[17] have proven to work well for histopathological tasks such as stain normalization[18,19] and stain-to-stain translation.[20] However, because of the complexity of the task, the precision of the pixel-wise loss functions[16] has been the rationale behind the use of supervised methods in most of the work thus far, on virtual staining of unstained tissue images.[21]

Recently, GAN-based image-to-image translation methods have been tested in studies using tissue specimens extracted from different organs, distinct types of input label-free tissue images obtained with various imaging modalities, and different stains to compare virtual staining against chemically stained tissue images.[21] Bayramoglu et al.[22] used conditional GAN and dimension reduction to virtually translate hyperspectral lung tissue label-free images to their H&E equivalent. Visual analysis showed that their method is promising; however, the quantitative assessment was inconclusive. Rivenson et al.[23] proposed a method called PhaseStain that used label-free quantitative phase images of human skin, kidney, and liver tissue to generate their virtual equivalent of H&E, Jones', and Masson's trichrome stain, respectively. Through quantitative evaluation, it was demonstrated that virtually stained tissue images were of high quality, and although the network output was sensitive to granular details (tested through phase noise) it was less affected by macro-level variability such as wrinkles and height variation pertaining to the tissue section.

Initial studies focused more on evaluating the image translation quality through image analysis quantitative evaluation metrics. More recent studies have also demonstrated the clinical potential of virtually stained images. Rana et al.[24] used a variant of conditional GAN[25] to virtually apply H&E stain to deparaffinized formalin-fixed paraffin-embedded prostate (core biopsy) tissue images and then de-stain the virtually stained images. In addition to direct quantitative evaluation, they used a tumor segmentation task to further test the clinical utility of their virtual staining method. The same group later built an end-to-end deep learning framework for the automatic detection and localization of tumors on the virtually stained images.[26] Some studies have even ventured into the experimental domain of novel virtual stains. Zhang et al.[27] proposed a novel solution for multiplex staining that digitally blends existing stains to generate new histological stains that are physically not possible yet. Along with stain blending, it also generated different micro-structured regions stained with H&E, Jones', and Masson's trichrome stain within the same tissue. In a recent study, we used CycleGAN,[17] an unsupervised learning approach, to determine the optimal unstained tissue processing and imaging protocols and showed proof of principle that deparaffinized and, in some cases, still-in-paraffin tissue sections could be used to achieve meaningful histology with virtual H&E staining.[28]

Although the aforementioned studies have explored different aspects of virtual histopathology staining including imaging technique, tissue types, and stains, to the best of our knowledge, the effect of the network capacity on the quality of the generated virtual histopathology images remains unknown. Furthermore, although pixel-wise quantitative metrics and general histological views have been presented, detailed histological accuracy of virtual H&E staining from standard brightfield images has not been reported thus far. Therefore, in this study, we systematically explored the effect of increased network capacity on the virtual staining of label-free, brightfield tissue images. In addition, to

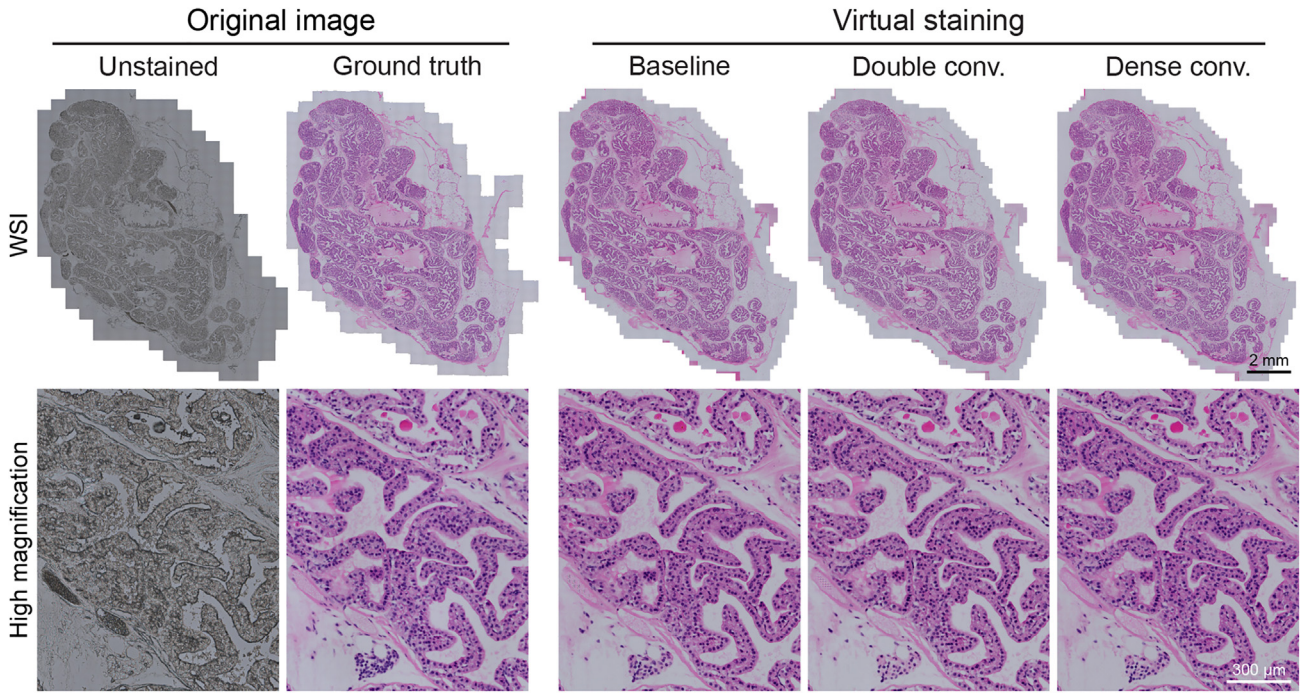


Figure 1. Overview of virtual staining performance

Example views of WSI and tissue level showing the performance of virtual staining. Each column shows original images of unstained or H&E-stained tissue, or virtually stained tissue with baseline, double convolution, or dense convolution architecture. Scale bars per row.

demonstrate the level of histological accuracy, we analyzed the results both quantitatively and qualitatively.

RESULTS

Structural similarity of virtual and chemical H&E staining on whole-slide level

We imaged unstained tissue sections of pre-clinical prostate tissue with standard brightfield microscopy and produced virtually stained H&E images with three variants of the supervised image-to-image translation model pix2pix.[25] To study the virtual H&E staining performance, we first visually compared virtually stained images with chemically H&E-stained ones from the same tissue sections at the whole-slide level. At a low magnification, all three variants, baseline, double convolution, and dense convolution, performed well, reproducing a highly similar macroscopic appearance of tissue histology as chemical H&E staining (Figure 1).

To further guide the visual evaluation, we used tile-wise structural similarity index measure (SSIM) scores to generate a heatmap for each whole-slide image (WSI) in the test set (Figure S1). The heatmap visualization revealed that a significant number of tissue edge tiles having only a small percentage of tissue content had high SSIM scores. This discovery led to the decision of including tiles containing only tissue content in the quantitative evaluation by excluding edge tiles containing white background as much as possible (Figure S2).

Quantitative evaluation of virtual H&E staining methods

Following visual inspection of the macro-level correspondence between virtual and chemical H&E staining, we proceed to quantitative evaluation of staining similarity for the three pix2pix variants using pixel-level similarity metrics. The first variant, a baseline pix2pix model that uses a single convolution layer on each level of the encoder and decoder, was trained to set a benchmark. It achieved a mean peak signal-to-noise ratio (PSNR) of 22.609, a mean SSIM of 0.725, and a mean Pearson correlation coefficient (PCC) of 0.903. The second variant, a pix2pix model with double convolution encoder-decoder blocks was trained to observe if it improved the quality of virtual staining over the benchmark training. The mean PSNR and SSIM did not improve, but PCC improved slightly; the values were 22.214, 0.720, and 0.904, respectively. Finally, the third variant, a pix2pix model with dense convolution encoder-decoder blocks, an approach inspired by DenseU-net,[29] was trained to explore the effect of a more sophisticated approach to increasing the network capacity on the quality of virtual staining. The dense convolution approach outperformed the previous two with a mean PSNR of 22.865, a mean SSIM of 0.746, and a mean PCC of 0.916. The overall sample-wise results can be seen in Table 1, the figures with asterisks represent the highest score. This trend holds even for evaluation that includes the edge tiles (Table S1). In addition to sample-wise mean scores, the results were further compared through violin and density plots of tile-level SSIM, PSNR, and PCC scores as shown in Figure 2, which shows a higher density of tiles toward high scores for the dense convolution approach as compared with the other two approaches in all three evaluation metrics.

Histological examination of virtual H&E staining quality

We then systematically evaluated the histological performance of the virtual H&E stainings produced by the three variants of

Table 1. Results from virtual staining experiment excluding edge tiles

	SSIM			PSNR			PCC		
WSI	Baseline	Double Conv.	Dense Conv.	Baseline	Double Conv.	Dense Conv.	Baseline	Double Conv.	Dense Conv.
Sample 1	0.733	0.738	0.719	21.718	21.845	21.704	0.853	0.868	0.869
Sample 2	0.734	0.733	0.729	21.481	21.516	21.604	0.853	0.866	0.868
Sample 3	0.716	0.725	0.726	21.514	21.800	21.957	0.874	0.885	0.890
Sample 4	0.757	0.764	0.764	22.597	22.981	23.205	0.883	0.894	0.900
Sample 5	0.712	0.722	0.724	21.717	22.021	22.039	0.872	0.883	0.887
Sample 6	0.648	0.649	0.680	20.520	19.922	20.083	0.922	0.928	0.934
Sample 7	0.699	0.675	0.729	22.034	20.555	21.465	0.923	0.912	0.930
Sample 8	0.773	0.769	0.800	24.518	24.121	24.652	0.929	0.933	0.940
Sample 9	0.772	0.753	0.798	24.251	23.659	24.839	0.928	0.919	0.933
Sample 10	0.716	0.704	0.758	23.671	22.716	23.915	0.928	0.921	0.940
Sample 11	0.735	0.737	0.778	23.527	23.130	24.430	0.921	0.919	0.934
Sample 12	0.708	0.692	0.745	23.263	22.551	23.919	0.929	0.917	0.940
Sample 13	0.715	0.696	0.752	23.110	21.959	23.437	0.926	0.909	0.938
Mean*	0.725	0.720	0.746*	22.609	22.214	22.865*	0.903	0.904	0.916*
SD*	0.033	0.035	0.039	1.216	1.170	1.688	0.031	0.022	0.0036

Tile-wise comparison of baseline, double convolution, and dense convolution experiments' virtually H&E-stained images against the chemically H&E-stained ground truth images. The numbers with asterisks represent the best results.
Conv., convolution; PCC, Pearson correlation coefficient; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

the pix2pix model with the prostate dataset. First, we evaluated the accuracy of different tissue components and types in the virtually stained images. The murine prostate consists of glandular structures lined with epithelial tissue abundant in eosinophilic secretion. The glands are surrounded by stroma consisting of connective tissue and smooth muscle in addition to the occasional adipose tissue, blood vessels, capillaries, and peripheral nerves.[30] In the three different virtual staining versions, all tissue components and types were readily detectable. Epithelial and adipose tissue stainings were reproduced particularly well (Figure 3). Tiles containing simpler structures, like adipose or secretions, were in fact the best-performing ones. Stroma, including muscle and connective tissue, nervous tissue, blood vessels, and capillaries appeared to be more challenging to re-create.

While inspecting cell and nuclear structures at high magnification, we observed that round nuclei in epithelium had accurate shape, size, and location (Figure 3). In other tissue types, most round and strongly basophilic nuclei were accurately reproduced. In each of the three architectures, the accuracy of nuclear representations decreased with elongated or irregular-shaped nuclei, or if the nuclei were hypochromatic with a particularly pale hematoxylin color tone (Figures 3 and S3). Eosinophilic areas performed well, and color tones in all tissues, particularly in muscle, nervous, and stromal loose connective tissue (Figure 3) were reproduced with high accuracy.

Because GAN-based image-to-image translation methods are notorious for hallucination artifacts, we screened for such in the virtually stained H&E images. Hallucination artifacts appeared in the baseline network-generated images, particularly where erythrocytes and lymphocytes were clustered together. However, these started to disappear when the network capacity was increased and seemed to completely vanish in the dense convolution experiment results (Figure S3).

Since the general representation of the tissue histology was astonishingly accurate, we tested the limits of histological accuracy that correlate with the potential clinical applicability of the methods in the future. We performed rigorous screening of all virtually stained images against the ground truth H&E images and mapped all possible histological representations requiring further attention in future model development. We found occasional fine patterning visible in virtual stainings from all three models (Figure S3); however, its source is not clearly evident. Other minor artifacts were misrepresented coloring, where sometimes the tissues would appear notably more or less blue or pink than in the ground truth; however, the histological interpretability was not affected by these (Figure S3). Occasionally, the edges of the tiles were visible, mostly in eosinophilic areas or outside of tissue (Figure S4). In some of the samples, a fine pattern can be detected in the output of all three network architectures (Figure S4). Although this pattern can be seen within the tissue, it is best visible outside of the tissue and does not impede histological interpretation.

Because histological sections sometimes have tissue artifacts, we also recorded interpretation of them by virtual H&E staining. These included patterns created by wrinkles and out-of-focus areas as well as crystal structures (Figure S5), the first one being by far the most common and a known challenge with chemically stained H&E. Interestingly, baseline network-generated images had the most distinct and largest pattern, resembling a fish-scale pattern, which decreased in size as the network capacity increased. Dense convolution did not have a defined pattern but a more indistinct blur of tissue-like structures. Sporadic black debris in the tissue was transferred correctly by the algorithm (Figure S5). Hematoxylin from H&E staining causes occasional debris in the chemically stained ground truth, which is interestingly eliminated by the use of virtual staining (Figure S5).

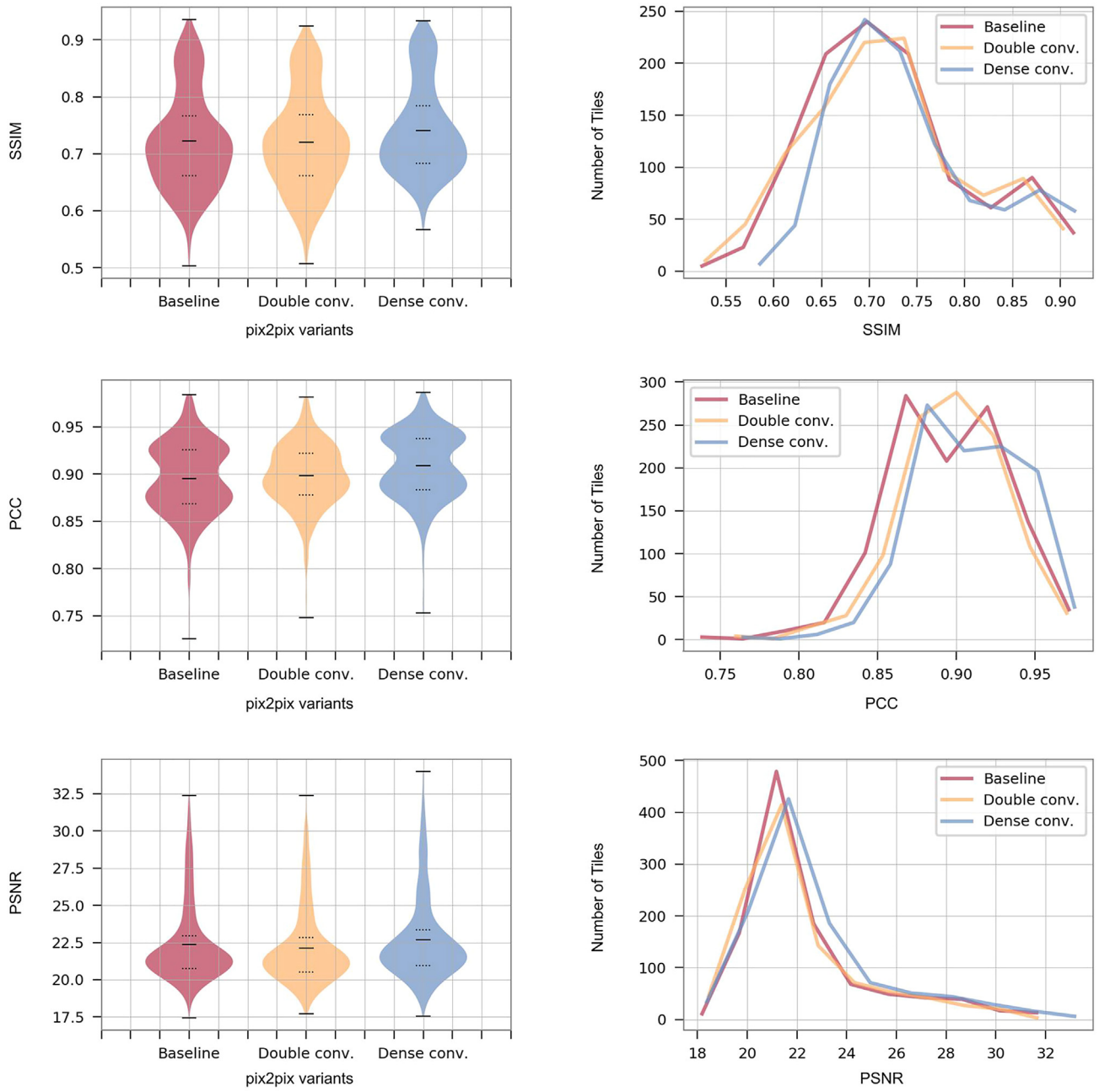


Figure 2. Quantitative evaluation plots

Tile-wise violin and density plots of SSIM, PCC, and PSNR scores for the three variants of the pix2pix model. Violin plots show mean, first quartile, minimum, and maximum values.

Representation of nuclear morphology in virtual H&E staining is of particular interest due to its importance in clinical pathology including cancer diagnostics. It appears that the more organized the epithelium is, the more accurately it is reproduced (Figure 4). Occasionally, nuclear fusions appear with adjacent nuclei in virtual staining. This phenomenon was detected in the output of all three architectures at a similar rate. On the other hand, there were differences in chromatism of nuclei between the algorithms, as falsely hyperchromatic nuclei appeared more abundantly with baseline than the other two architectures. As opposed to the well-performing epithelium with nuclei more regular in shape and location, elongated nuclei were challenging to virtually reproduce. This included mainly basal cells whose location or shape was incorrect, and in some cases, false positive nuclei appeared. Although a lot of variation in fine texturing exists in the virtually stained H&E images, in some nuclei even intranuclear details, such as heterochromatin, were correctly interpreted (Figure 4). Further, even

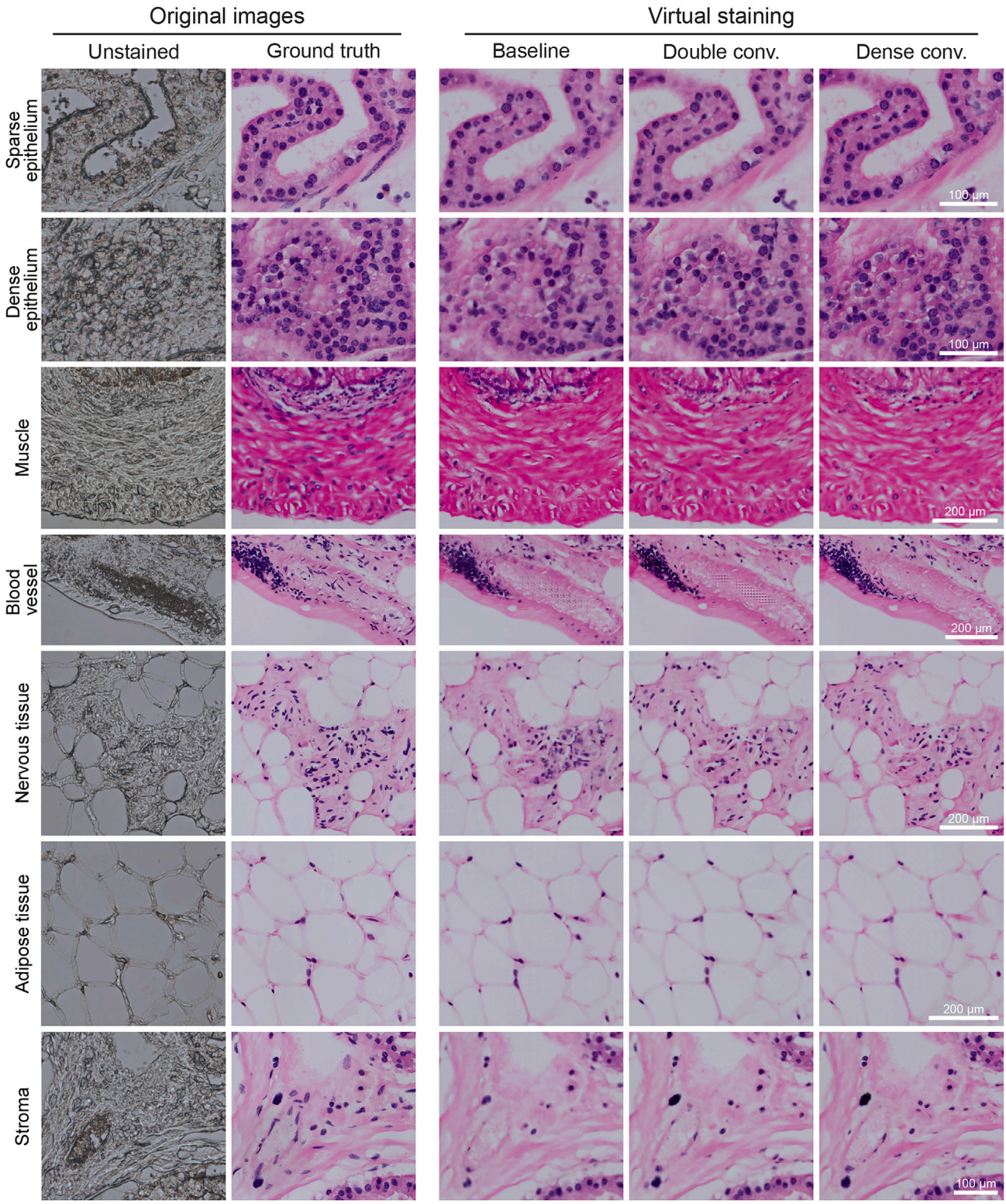


Figure 3. Representative images of different virtually H&E-stained tissue types and the corresponding unstained tissue and ground truth

Each column shows original images of unstained or H&E-stained prostate tissue, or virtually stained tissue with baseline, double convolution, or dense convolution architecture. Scale bars per row.

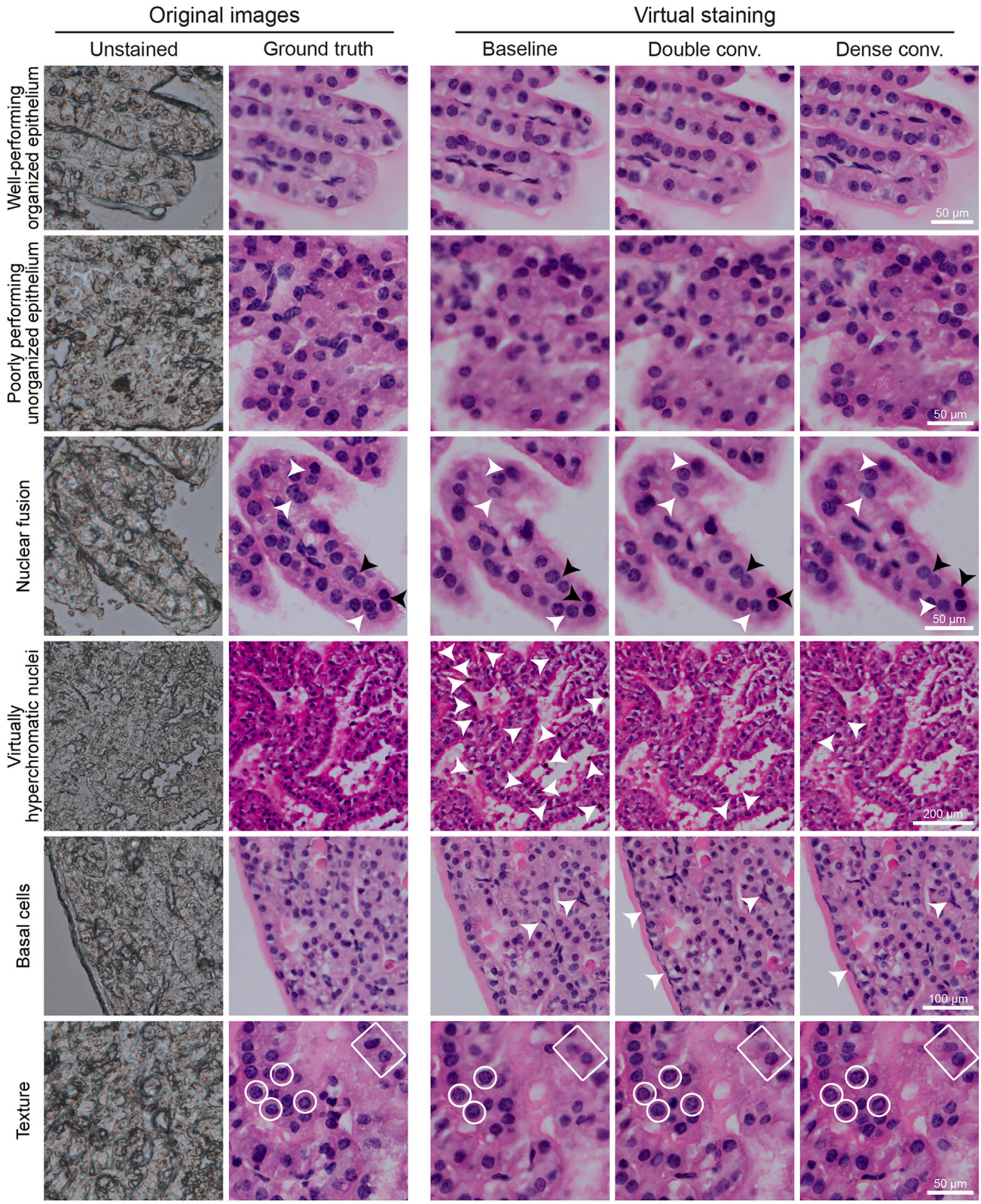


Figure 4. Representative images of nuclei in virtually H&E-stained tissue images

Images show nuclear performance in prostate epithelium with the virtual staining algorithms. Virtual staining algorithms generate more accurate nuclei representations in the organized epithelium compared with unorganized epithelium (two upper rows). Infrequent nuclear fusions (third row, white arrows) occur

(legend continued on next page)

cell membranes could also be distinguished in certain eosinophilic regions (Figure 4).

Quantitative evaluation of nucleus-level statistics from virtual stained images

In addition to the WSI level similarity using pixel-level metrics, we evaluated the resemblance between H&E and virtual staining quantitatively on a single-cell level by comparing nucleus detection statistics for the prostate dataset. A nucleus segmentation model called Hover-Net[31] was used for segmenting nuclei in patches extracted from H&E-stained reference images, and similar segmentation was performed to virtually stained counterparts. The original Hover-Net model was used with the CoNSeP checkpoint. As an output, the model provides centroid coordinates and contours for each nucleus. In total, 135 image tiles of 2048 × 2048 pixels (723 × 723 μm) were extracted from H&E reference stained samples, covering all WSIs in the test set of the prostate dataset, and including areas of various tissue types (epithelial, stromal, adipose, muscle). The statistics reported for the nucleus segmentation accuracy are shown in Figure 5, illustrating overall nucleus counts per tile, and object-level statistics averaged per tile as F-score, precision, and recall. For calculating the nucleus-level correspondence, a true positive is defined as the centroid coordinate of nucleus segmentation from H&E hitting the nucleus segmentation mask for a virtually stained image. A false negative is an H&E nucleus coordinate landing on the background (non-nucleus) in segmentation of virtually stained, and a false positive is a virtually stained segmentation without any matching H&E nuclei coordinates within the nucleus mask.

Example tiles with nucleus segmentation overlaid are shown in Figure 5A for H&E, baseline, double convolution, and dense convolution. The correlation between nucleus counts from H&E and virtually stained tiles show very good correspondence (Figure 5B). All virtual staining methods reproduce nuclei in very similar numbers compared with ground truth H&E as estimated by the nucleus segmentation model, with PCC of tile-level counts ranging between 0.912 and 0.952. Object-level correspondence measured as F-score shown in Figure 5C, however, is suboptimal. Despite producing roughly the same quantity of nuclei in visually similar location patterns, their locations do not fully match, leading to relatively low correspondence when measured by F-score for all three methods. However, also based on the F-scores, the dense convolution variant outperforms other virtual staining methods, mainly due to its higher precision when compared with baseline and double convolution (Figure 5C). The results indicate high accuracy of recognition of the virtual staining-produced nuclei and support the results of the histological evaluation in that nuclei are relatively well reproduced especially in epithelial compartments, while there is a suboptimal representation of elongated, e.g., stromal nuclei with the current solution.

Applicability of virtual H&E staining on multiple tissues

As our virtual staining algorithm showed promising results and accuracy with anterior prostate tissue, we wanted to further explore the applicability of the virtual staining method with other organs with different morphologies. Our tissue panel consisted of three relatively homogeneous tissues in their composition, namely liver, spleen, and kidney, as well as three glandular tissues with distinct characteristics from the prostate, namely seminal vesicle, testis, and epididymis. We conducted virtual staining with the dense convolution model, which performed best with the prostate dataset. Comparing the virtually stained images to chemically stained H&E ones from the same tissue sections reveals that, at the WSI level, virtual stainings of all the six tissues have representative morphology, and that with higher magnifications, most of the distinct structures of each organ are clearly distinguishable (Figures 6A–6F).

Liver tissue consists of hepatocytes and has special characteristics with portal regions.[32] Both of these distinctive features are well characterized in the virtual staining (Figure 6A). Seminal vesicles are surrounded by a thin layer of stroma-like connective tissue, lined with branched tall columnar epithelium and filled with intensely eosinophilic secretion.[30] The epithelial tissue and secretion are well reproduced in the virtual staining, while the connective tissue lining is frequently missing the elongated nuclei (Figure 6B), similar to the prostate tissue results (Figure 3). The spleen is a nuclei-dense organ that is grossly divided into two zones: the erythrocyte-rich red pulp and the basophilic white pulp,[33] both of which can be observed in the virtually stained spleen (Figure 6C). The kidney has two distinct regions, the outer cortex and the inner medulla, both consisting largely of cuboidal epithelium,[34] which performs particularly well in virtual staining (Figure 6D). The glomeruli, which are largely irregular in shape, can be distinguished but their more nuclei-dense areas were less well reproduced. The testis is composed primarily of convoluted seminiferous tubules,[30] the structure of which can be observed clearly in the virtually stained images (Figure 6E). The tubules of the testis contain the maturing spermatogenic cells, which are basophilic in nature,[30] and although most of them are visible in the virtually stained images, some have not been reproduced (Figure 6E). The epididymis comprises epididymal ducts, which are lined with thin epithelium and the lumen is filled with mature spermatozoa.[30] The epithelial structures as well as the lumen are represented well in the virtually stained tissue, although there is occasional fading of epithelial structures and fusion of overrepresented nuclei (Figure 6F).

DISCUSSION

We present here the first work to assess the detailed histological feasibility and the effect of neural network architecture on virtual H&E staining from unstained tissue images using GANs. In this study, we compared three different variants of one of the most commonly used supervised image-to-image translation methods called pix2pix on the task of virtual H&E staining of deparaffinized unstained tissue images. The three variants were the

between adjacent nuclei. Black arrows indicate correctly represented adjacent nuclei. Hyperchromatic nuclei are created by virtual staining, especially with the baseline architecture (fourth row, white arrows). Occasional false negative basal cell nuclei appear with all three architectures (fifth row, white arrows). Correctly interpreted fine textures (cellular structures indicated with rectangles, nuclear texture indicated with circles) are shown in the bottom row. Each column shows original images of unstained or H&E-stained tissue, or virtually stained tissue with baseline, convolution, or convolution. Scale bars per row.

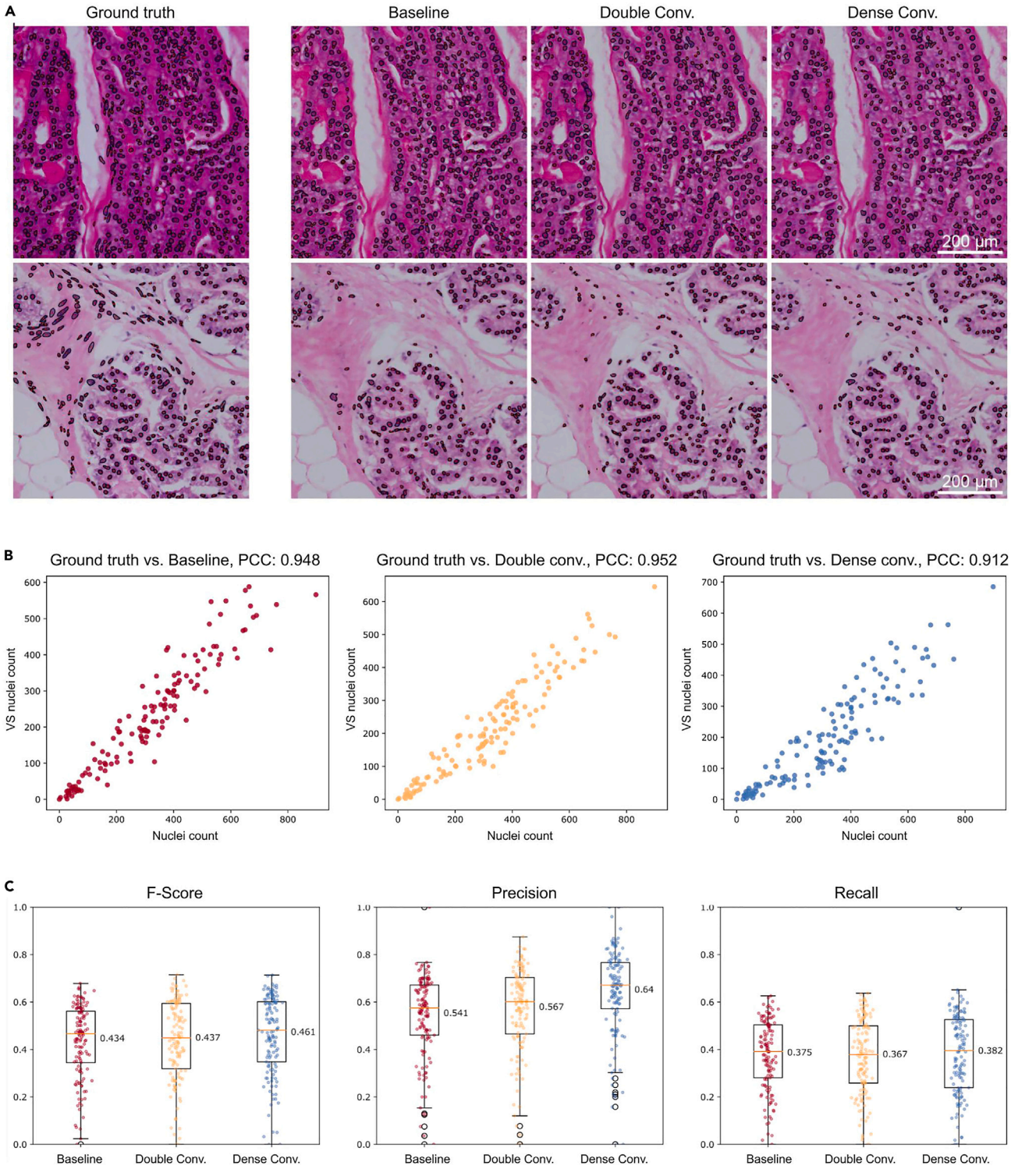


Figure 5. Nucleus-level quantification accuracy in virtually stained H&E images

The performance of Hover-Net[31] nucleus segmentation for virtually stained images was quantitatively evaluated by using segmentation obtained for the H&E staining as the ground truth.

(A) Examples of nuclei segmentation masks based on H&E-stained tissue, and virtually stained tissue with baseline, double convolution, and dense convolution architecture. Scale bars per row.

(legend continued on next page)

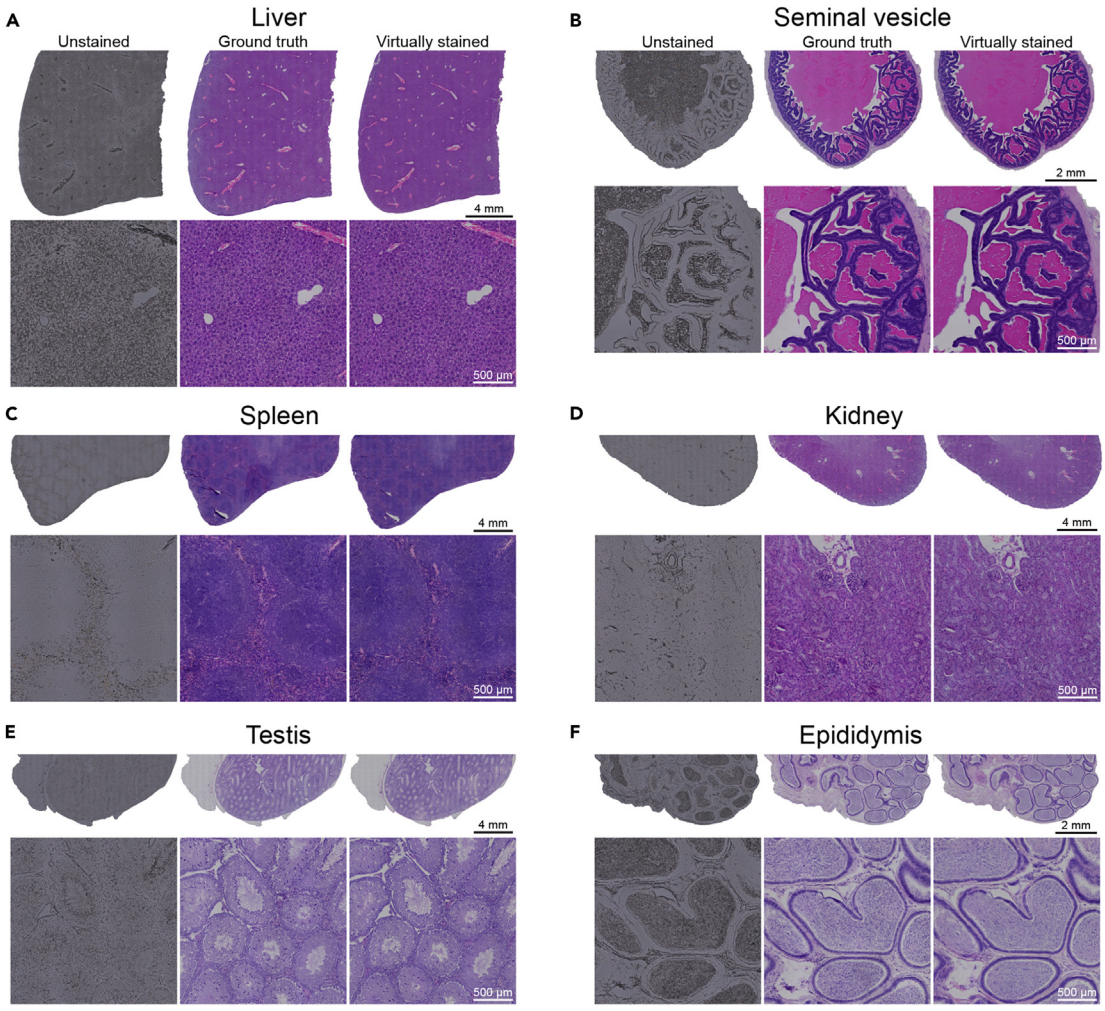


Figure 6. Virtual H&E staining of different organs

(A–F) Tissue images representing different organs virtually stained with dense convolution variant of pix2pix. Images show WSIs at a low magnification (upper rows) and areas of high magnification (bottom rows) displaying histology of (A) liver, (B) seminal vesicle, (C) spleen, (D) kidney, (E) testis, and (F) epididymis. Scale bars per row.

baseline, which is the original implementation of pix2pix, the double convolution in which convolution layers were doubled in the encoder and the decoder of the generator, and the dense convolution in which DenseU-net[29] style encoder and decoder were used in the generator. Doubling the convolution operations per layer was an intuitive first modification to the baseline, which was expected to produce better results by honing the features per layer of the generator. It is a common approach in modern convolutional neural networks like InceptionV3[35] and ResNet[36] to use multiple convolution operations before pooling the intermediate output. The motivation behind the use of a DenseU-net-inspired approach was its success in small object detection in satellite imaging. The idea was to test whether the approach could be used for virtual staining to accurately reproduce smaller and finer details like the shape of nuclei. The purpose of these modifications was to understand the effect of increased network capacity on the visual quality of virtual staining. Quantitative evaluation metrics and rigorous histological analysis for morphological accuracy were used to establish that, although all networks produced noteworthy H&E virtual stainings, dense convolution virtual staining quality was superior to both baseline and double convolution. The inference time of dense convolution is approximately 1.5X and 2X as compared with double convolution and baseline, respectively. Even with dense convolution, it takes slightly more than 2 min on the hardware setup used in our study to virtually stain a WSI, which is significantly faster than the chemical staining process and reasonable for practical applicability.

(B) Nuclei count correlations between ground truth H&E staining and pix2pix variant virtual stainings. Each data point corresponds to the number of nuclei detected in a tile.

(C) Tile-level averages of nucleus detection F-score, precision, and recall for virtual stainings against the segmentation result obtained for the H&E-stained ground truth. Boxplots show the mean, first and third quartiles, and 1.5x interquartile range as whiskers.

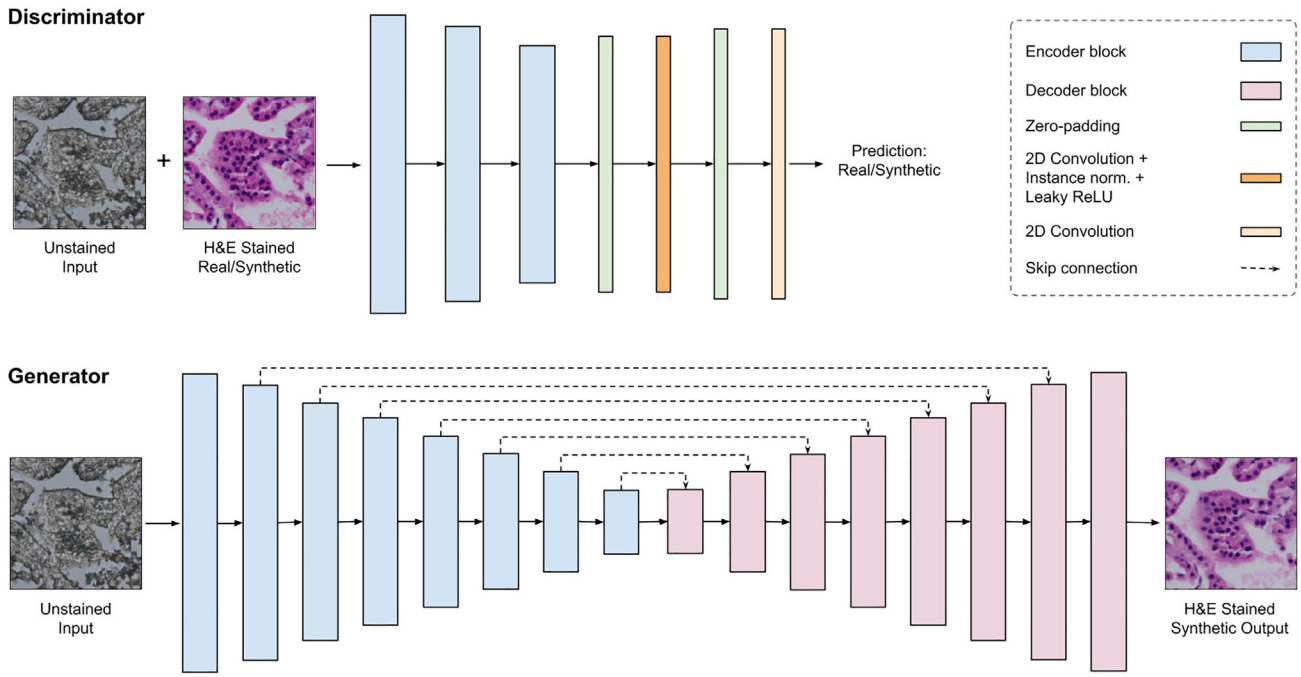


Figure 7. The pix2pix model architecture

The PatchGAN discriminator comprises three encoding blocks, the first one without instance normalization, followed by a zero-padding layer, a 2D convolution layer, instance normalization, a leaky rectified linear unit activation, another zero-padding layer, and finally a 2D convolution layer. The U-net generator consists of eight encoding and eight decoding blocks. Each encoder block further consists of a 2D convolution layer with a stride size of two, followed by instance normalization and leaky rectified linear unit activation. Each decoder block consists of a 2D transposed convolution, with a stride size of two, followed by instance normalization, and leaky rectified linear unit activation. Skip connections were used in the generator, which means that the output of each encoding block is concatenated with the output of the corresponding decoding block excluding the first encoding and last decoding block.

Pathological evaluation in the clinics is currently dependent on chemical H&E stainings used for visual assessment to, e.g., diagnose cancer. The possibility of avoiding chemical staining would enhance workflows and enable cost savings through decreased laboratory work and chemical consumption. The utility of virtual stainings, however, depends on the use cases. To replace the chemical H&E staining in visual assessment tasks, the accuracy of the virtual staining needs to reach the level enabling reliable visual interpretations. Hence, we thoroughly screened the virtual stainings for their microscopic histological performance and to identify emphasis points for future method development.

We found that the morphology of the tissues at both micro- and macroscopic levels was highly accurate. In the three different virtual stainings, all tissue components and types were readily detectable. We first used murine prostate tissue, which is abundant in epithelial tissue, but also contains all the other major tissue types. This enabled us to get a general overview of the tissue types and their attributes that are reproduced well by the algorithms and the ones that appear more challenging. Epithelial and adipose tissue performed particularly well, and epithelial nuclei were well reproduced. Upon high magnification, those tissue and cell types that were not as abundantly represented in the training data, such as nervous tissue, vessels, or basal cells, had features that were more difficult to re-create. This is a typical challenge in deep learning, one that is likely to be solved by increasing the amount of data for these specific tissue components in the training.

In virtual staining, eosinophilic areas performed exceptionally well, for which color tones in all tissues were interpreted with high accuracy with all three virtual staining architectures. Virtual staining even had benefits over chemical H&E staining in clearing debris left behind from hematoxylin. With successfully chemically stained H&E, even various intracellular details can be distinguished from their staining intensity, texture, and color tones.[37] We found that the higher capacity networks produce high-accuracy H&E images compared with ground truth and that even subcellular accuracy is often high. However, interpreting the details such as sub-nuclear morphology should still be done with caution as, depending on the architecture and area, there is room for improvement especially with color toning and finer details.

Increased network capacity clearly benefited the histological accuracy of virtual staining. This was evidenced by a reduction in hallucination artifacts, tissue section artifact-created patterns, and virtually hyperchromatic nuclei. The baseline pix2pix performance was histologically inferior to double convolution and dense convolution in most tissue types and, overall, the dense convolution performed best. Double convolution had more false positive nuclei than the other two architectures, potentially contributing to the lower average SSIM and PSNR scores. Overall, the higher SSIM and PSNR scores did not always correlate

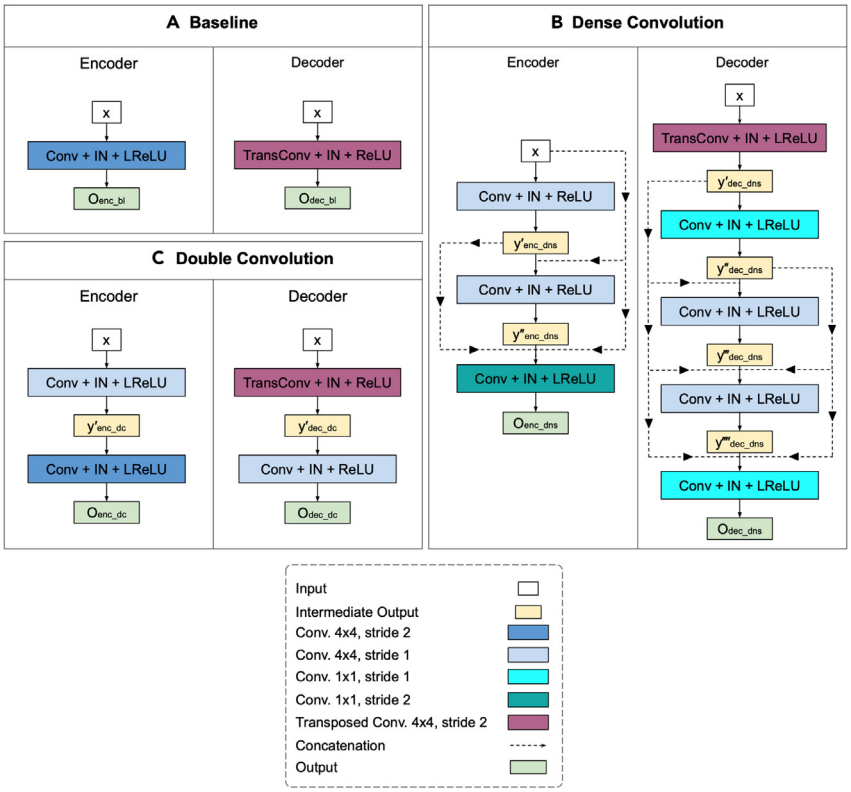


Figure 8. Three variants of pix2pix encoder and decoder blocks

(A) Baseline: the reference pix2pix implementation with single convolution for each encoder and decoder block.

(B) Double convolution: additional convolutional layer for each encoder and decoder block.

(C) Dense convolution: a more complex unit inspired by DenseU-net[29] (used for the segmentation of small objects in remote sensing images) for each encoder and decoder block.

fully with the visual inspection, and the tissue composition of each WSI and the amount of background included in the calculation substantially influenced the scores. Thus, in future assessments, tissue type-specific scores could give an enhanced numerical interpretation of the results and improve relevance to diverse pathologies manifesting in different tissue types.

After studying the effect of network capacity on virtual H&E staining of prostate tissue, we tested the applicability of the best-performing network to virtually stain tissue images from six other organs. We saw a similar performance to prostate tissue virtual staining, where the overall morphology at the WSI level appeared indistinguishable from chemical staining. The more homogeneously organized tissues were well interpreted, as were the epithelial tissues with morphological resemblance to the prostate. With the tissue panel dataset, we noted similar artifacts, such as patterning in more nuclei-dense areas and out-of-focus regions, missing elongated nuclei as well as occasional nuclear fusion (data not shown) as with the prostate dataset. Our results demonstrate that the dense convolution model performs well with several tissues and can likely be applied to additional tissues as well. Overall, the accuracy demonstrated here for virtual H&E staining with pix2pix and dense convolution is sufficient for several histological purposes while bearing in mind the limitations of the method in detailed histology.

Even though we show that increasing network capacity can in fact improve the quality of virtual staining, it is purely an algorithmic aspect and there are other factors that affect the overall quality of virtual staining. For instance, the H&E color manifestation varies from batch to batch and from laboratory to laboratory (Figure S6). Stain normalization methods were not used in this study to keep tissue images as unaltered as possible. In the future, both stained and unstained tissue images could be normalized beforehand to explore the impact of color uniformity on virtual staining.

Most prior reports on virtual staining use specific imaging or spectroscopic modalities, such as Fourier transform infrared spectroscopy,[38] quantitative phase imaging,[23] or fluorescence.[27,39] These require specific instruments not readily available in most clinical settings. Hence, we wanted to develop approaches using standard brightfield imaging for which the same microscopes and slide scanners could potentially be used as with chemically stained H&E slides.[28] Virtual staining methods working on the standard platform images have vast potential to streamline histopathological workflow without requiring additional equipment or other investments. Prior studies[24,26,28,40] using brightfield images for virtual H&E staining reported histology mainly at the macroscopic level. We used imaging with 40× magnification producing high-level reproduction of histology at a microscopic level. Although images obtained with 20× magnification have been reported to suffice for clinical tumor identification,[24] future work is needed to assess the potential and requirements of brightfield imaging setup and image resolution for histological utility in different pathological tasks.

This study highlights the advantages and challenges of virtually staining unstained tissues and demonstrates how increasing neural network capacity improves the performance of virtual staining both at the quantitative and visual levels. To study the clinical significance of increased neural network capacity, diagnostic accuracy could be compared by visual observation between chemically and virtually stained H&E in future experiments using tumor tissue. Furthermore, similar to Bayat et al.,[26] another tumor segmentation network could be used to compare the segmentation results of virtually stained tissue images against chemically stained. Here, we showed good concordance in nuclei counts obtained using a dedicated deep learning tool,[31] showing promise for the use of virtual staining of unstained tissues in performing various tasks based on nuclear counts,[41] and explored the limitations in reproducing nuclei in the exact same locations. With the encouraging results so far from computationally generating information using

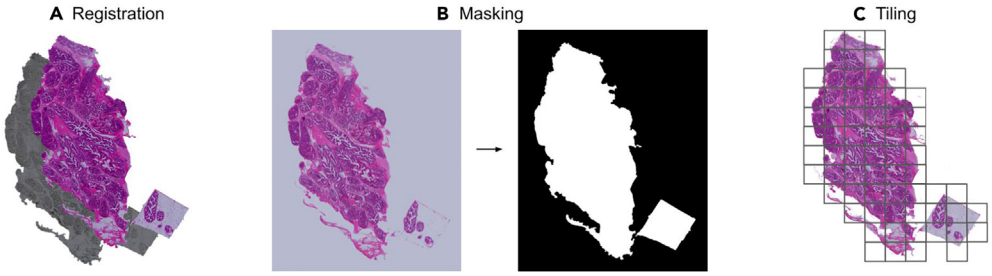


Figure 9. Preprocessing of stained and unstained tissue images

(A) Registration: Unstained and stained images are first rigidly aligned followed by elastic registration.

(B) Masking: Binary masks for tissue regions in both unstained and stained images are generated.

(C) Tiling: The masks are used to generate tiles only from regions that contain tissue.

unstained tissue sections that is currently gained through H&E chemical staining, virtual staining shows great promise in streamlining sample processing for specific use cases in future pathology.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources should be directed to the lead contact, Pekka Ruusuvuori (pekka.ruusuvuori@utu.fi).

Materials availability

This study did not generate any new unique materials.

Data and code availability

- The tissue whole-slide image dataset used for quantitative evaluation in this study is freely available on a FAIR-compliant server under the https://doi.org/10.23729/9ddc2fc5-9bdb-404c-be07-c9c9540a32de. Refer to Table S2 for test set sample names lookup table.
- The implementation of all three variants of the pix2pix model, used in this study, has been deposited at Zenodo under the https://doi.org/10.5281/zenodo.7589356 and are publicly available as of the date of publication. Refer to *readme.md* for detailed instructions on how to use the code and *requirements.txt* to install the dependencies.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Methods exploration

Virtual staining is an image-to-image translation problem for which there exist both unsupervised and supervised learning-based approaches and their combinations in semi-supervised methods. To explore the potential of the different learning paradigms and their different implementations, we conducted experiments with several image-to-image translation methods covering a variety of approaches. First, two different variants of CycleGAN,[17] a commonly used unsupervised method, were used for virtual staining. In the first variant, generators were ResNet-inspired[36] and in the second one U-net-inspired.[12] The first variant was also modified to devise a semi-supervised approach, along with unpaired image data, it was also trained with varying percentages of paired data. For these batches, an L1 loss function was used instead of the cycle consistency loss (Figure S7). Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation (U-GAT-IT),[42] another unsupervised method, was also screened for the task of virtual staining (Figure S8). Based on visual analysis, the second CycleGAN variant, with U-net-based generators, performed relatively better than other unsupervised methods (Figure S9).

Using unsupervised and semi-supervised approaches first was driven by the motivation to make the training process more efficient by having little to no dependence on image registration. Unfortunately, all the above experiments produced suboptimal results, which were easy to discern even by visual analysis, and hence were not deemed suitable for further virtual staining experimentation. Methods such as one-shot[43] and few-shots training[44] were also tested out to no avail (Figure S10). The lack of success with the aforementioned methods in terms of visual quality as compared with pix2pix results during the initial exploration phase led to the decision of selecting pix2pix for further exploration and modification.

Pix2pix variants

Our implementation of the pix2pix model follows that of the work by Isola et al.[25] The baseline model comprised a generator based on U-net[12] and PatchGAN discriminator,[25] both components being fully convolutional neural networks as shown in Figure 7. The encoder and the decoder block output in the baseline variant are represented as follows:

$$O_{enc\_bl} = \sigma(W.x)$$
$$O_{dec\_bl} = \sigma(W^T.x)$$

where x is the input, O_{enc_bl} and O_{dec_bl} are the encoder and decoder blocks output, respectively. σ is the combined effect of normalization and activation. W and W^T represent the convolution and transposed convolution operations, respectively.

In the subsequent experiments, the encoder and decoder blocks of the generator were enhanced and its impact on the quality of virtually stained tissue images was studied. First, the convolution layers in each block of both the encoder and the decoder in the generator were doubled. Convolutions were also doubled in the encoding blocks of the discriminator, but it was observed that doubling the convolution layers in the discriminator makes it too dominant too early in the training hampering the learning of the generator. Since the generator unit is the actual image-to-image translation module, the PatchGAN discriminator was used as is in all the experiments. The double convolution encoder block has an intermediate output which is represented as follows:

$$y'_{enc\_dc} = \sigma(W_1.x)$$

where y'_{enc_dc} is the intermediate output, σ is the combined effect of normalization and activation, and W_1 represents the first convolution operation. The encoder final output is represented as follows:

$$O_{enc\_dc} = \sigma(W_2.\sigma(W_1.x))$$
$$O_{enc\_dc} = \sigma\left(W_2.y'_{enc\_dc}\right)$$

where O_{enc_dc} is the final output of the decoder block and W_2 represents the second convolution operation. Similarly, the double convolution decoder block also has an intermediate output, which is represented as follows:

$$y'_{dec\_dc} = \sigma(W^T.x)$$

where y'_{dec_dc} is the intermediate output, σ is the combined effect of normalization and activation and W^T represents the transposed convolution operation. The decoder final output is represented as follows:

$$O_{dec\_dc} = \sigma(W.\sigma(W^T.x))$$
$$O_{dec\_dc} = \sigma\left(W.y'_{dec\_dc}\right)$$

where O_{dec_dc} is the final output of the decoder block and W represents the convolution operation.

Next, a more sophisticated approach was used to bulk up the generator network even further. The idea was to make each encoding and decoding block denser. To that end, the dense convolution unit approach used in the segmentation of small objects in remote sensing images[29] was adopted. Dense convolution units were developed for both the encoder and the decoder. The encoder block in the dense convolution variant has two intermediate outputs, which are represented as follows:

$$y'_{enc\_dns} = \sigma(W_1.x)$$
$$y''_{enc\_dns} = \sigma\left(W_2.\left(x + y'_{enc\_dns}\right)\right)$$

where x is the input to the encoder block. y'_{enc_dns} and y''_{enc_dns} are the first and second intermediate outputs. σ is the combined effect of normalization and activation function. W_1 and W_2 are first and second convolution operations, respectively. The final output of the encoder block is as follows:

$$O_{enc\_dns} = \sigma\left(W_3.\sigma\left(y''_{enc\_dns} + y'_{enc\_dns} + x\right)\right)$$

where O_{enc_dns} is the final output of the encoder block and W_3 represents the third and final convolution operation. The decoder block of the dense convolution variant has four intermediate outputs that are represented as follows:

$$y'_{dec\_dns} = \sigma(W^T.x)$$
$$y''_{dec\_dns} = \sigma\left(W_1.y'_{dec\_dns}\right)$$
$$y'''_{dec\_dns} = \sigma\left(W_2.\left(y''_{dec\_dns} + y'_{dec\_dns}\right)\right)$$
$$y''''_{dec\_dns} = \sigma\left(W_3.\left(y'''_{dec\_dns} + y''_{dec\_dns} + y'_{dec\_dns}\right)\right)$$

where y'_{dec_dns}, y''_{dec_dns}, y'''_{dec_dns}, and y''''_{dec_dns} are the four intermediate outputs. σ is the combined effect of normalization and activation function. W^T, W_1, W_2, and W_3 are the transposed, first, second, and third convolution operations, respectively. And the final output of the decoder is as follows:

$$O_{dec\_dns} = \sigma\left(W_4.\left(y''''_{dec\_dns} + y'''_{dec\_dns} + y''_{dec\_dns} + y'_{dec\_dns}\right)\right)$$

where O_{dec_dns} is the final output of the encoder block and W_4 represents the fourth and final convolution operation.

To summarize, three variants of the pix2pix model were implemented to virtually stain the unstained tissue images. They were baseline pix2pix, pix2pix with double convolution encoder-decoder blocks, and pix2pix with dense convolution encoder-decoder blocks. A schematic representation of the three encoder-decoder block variants can be seen in Figure 8.

Loss function

In all three variants of pix2pix, i.e., baseline, double convolution, and dense convolution, the same conditional GAN loss function was used. It is defined as follows:

$$L_{cGAN}(G,D) = E_{x,y}[logD(x,y)] + E_x[log\ (1 - D(x,G(x)))X]$$

where G and D represent the generator and discriminator, respectively. x is the input and y is the ground truth and $G(x)$ is the synthetic output. Here G tries to minimize the loss against an adversarial D that tries to maximize it, i.e.,

$$G^* = \underbrace{argmin}_{G}\ \underbrace{max}_{D} L_{cGAN}(G,D)$$

This was further extended by a weighted L1 loss term as suggested in the original paper to minimize blurriness of the generator output.

$$L_{L1}(G) = E_{x,y}\left[\|y - G(x)\|_1\right]$$

The final objective is represented as follows:

$$G^* = \underbrace{argmin}_{G}\ \underbrace{max}_{D} L_{cGAN}(G,D) + \lambda L_{L1}(G)$$

Tissue material

Murine tissues used in the study were surplus tissue from prior studies.[45,46] The prostate tissues were fixed in PAXgeneTM (PreAnalytiX GmbH, Hombrechtikon, Switzerland), whereas kidney, liver, spleen, testis, epididymis, and seminal vesicle tissues were fixed in formalin. This was followed by paraffin embedding. The tissues were sectioned to 5μm thickness, placed on slides, and attached in +37°C for 30 min. The paraffin was chemically removed by xylene wash, followed by rehydration ethanol washes, and finally a wash in distilled water. The slides were then air dried and imaged as WSIs unstained with Thunder Imager 3D Tissue slide scanner (Leica Microsystems, Wetzlar, Germany) equipped with DMC2900 camera and HC PL APO 40x/0.95 DRY objective with a pixel resolution of 0.353 μm. After imaging as unstained, the samples were stained with H&E by first rehydrating in distilled water, followed by Delafield's hematoxylin (1159380100, Merck, Darmstadt, Germany), running tap water, 120 mM HCl in 70% ethanol, running tap water, and eosin (1159350025, Merck). The staining was followed by standard dehydration by 96%, 100% ethanol, and xylene washes, after which the slides were mounted with coverslips and imaged again. The data consist of 81 WSI pairs of unstained and H&E-stained histological sections of anterior prostate tissue (prostate dataset), and one WSI pair of unstained and H&E-stained tissues per tissue type for the other tissues (tissue panel dataset).

Image processing

Image registration

Three-phase registration step was applied to the WSI pairs: First, a subset of WSI pairs was roughly registered using rigid registration, which uses translation, rotation, and scaling operations for image alignment. This subset of WSI pairs was used to train a baseline model to generate an intermediate H&E-like output for all the unstained WSIs. Since the training WSI pairs were rigidly aligned, the resulting virtual staining quality was suboptimal, but nonetheless, the intermediate output was imperative for the next step because of its inherent alignment with the input unstained WSI. Then, the ground truth H&E WSIs were elastically registered to the intermediate H&E-like WSIs as a proxy for unstained WSIs, because of similar appearances the resulting alignment was precise. This in turn refined the alignment between previously rigidly registered unstained and ground truth H&E WSI pairs. Elastic registration is a nonlinear content alignment technique that takes into account nonlinear sources of misalignment such as spherical distortions and nonuniform morphological changes in the tissue due to different chemical processes the tissue goes through. Dice scores for the binary masks of the image pairs were computed before and after registration. The mean dice score improved by approximately 9% from 0.89 to 0.98 after registration.

Masking

In a WSI, only regions containing the tissue content are important for the training of a neural network. The binary masks for the tissue content were generated using the following series of steps: grayscale conversion of the WSI, thresholding to remove background, dilation, binary hole filling, and then erosion in the same order. Only regions where tissue content was present in the registered unstained and stained WSI pair were used for further processing; this was done by generating an intersection mask of the WSI pair masks (Figure S11).

Tiling

The tissue masks were used to guide the tiling processing to only include tiles that contain the tissue content. Tiles of 512 × 512 pixels (180 × 180 μm) were extracted for the training process. Since all the models were fully

convolutional, the tile size was increased to 2048 × 2048 pixels (723 × 723 μm) for inference, this has been shown to generate output tiles with more consistent colors.[19]

Figure 9 illustrates an example of the above-mentioned preprocessing steps.

Virtual staining experiment setup

Virtual staining for the prostate dataset

The data were divided into two sets based on the visual appearance of the H&E-stained WSI. The data originated from two batches, where the color manifestation in one set was slightly darker than the other (Figure S6). Instead of applying stain normalization, we opted to keep the tissue images unaltered to avoid artifacts caused by the normalization methods, such as tiling artifacts on the WSI level,[19] or hallucination artifacts common in GAN-based methods,[16] in particular CycleGAN.[17]

The darker set contained 49 images; 39 for training, 2 for validation, and 8 for testing. The lighter batch contained 32 images; 25 for training, 2 for validation, and 5 for testing. Altogether, 64 slides were used for training, 4 slides for validation, and 13 for testing. A total of 1,149,516 tiles were used in the training process, and 20k tiles were randomly chosen for samples with more than 20k tiles. All the models were trained for 40 epochs. For each epoch of the network training, 50% of the training tiles were randomly rotated, flipped, and scaled. Epochs with the lowest validation loss were chosen for inference. Trainings were parallelized over four NVIDIA Volta V100 GPUs.

For the darker batch, the 40 epochs training of baseline, double convolution, and dense convolution took approximately 133 h, 213 h, and 340 h, respectively. For the lighter batch, the 40 epochs training of baseline, double convolution, and dense convolution took approximately 97 h, 160 h, and 230 h, respectively. For the lighter batch, the inference with baseline, double convolution, and dense convolution took approximately 7.5 min, 11.5 min, and 15 min, respectively. For the lighter batch, the inference with baseline, double convolution, and dense convolution took approximately 6 min, 9.5 min, and 14 min, respectively.

Virtual staining for the tissue panel dataset

The experiment setup for virtual staining for tissue panel data was performed largely as above with a few modifications. Tissue samples were halved, and the top part was used for training and the bottom part for testing. Tiles were sampled based on the size of the tissue, e.g., 35k tiles were sampled from liver tissue WSI for training, whereas, 9k were from seminal vesicles.

Evaluation metrics

The virtually stained tissue images were evaluated against their corresponding chemically H&E-stained ground truth images using three different evaluation metrics: SSIM,[47] PSNR,[48] and PCC.

SSIM is as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where x and y are the virtually stained image and the corresponding chemically H&E-stained ground truth image, respectively. μ_x, μ_y, and σ_x, σ_y are the mean and the standard deviation of images x and y, respectively. σ_{xy} is the covariance of images x and y, and c_1, c_2 are stabilization constants used to prevent division by a small denominator. The value of SSIM ranges from 0 to 1, the higher the value the more similar the images.

PSNR is:

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

where MAX_I is the maximum pixel intensity and MSE is the mean squared error between the two images. PSNR is commonly used to quantitatively evaluate the reconstruction quality of images and videos. The accurate value is considered to range between 20 dB and 25 dB and a higher PSNR is better.

PCC is:

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

where x and y are the virtually stained image and the corresponding chemically H&E-stained ground truth image, respectively. $cov(x,y)$ is the covariance of x and y images. σ_x and σ_y are the standard deviations of each image x and y, respectively. The value of PCC ranges from 0 to 1, the higher the value the more similar the images.

Histological evaluation of performance

While the computational evaluation metrics are suited for the overall comparison of network performance, the usability of the virtual stainings for visual pathological evaluation depends on the capacity to reproduce histological structures accurately. Hence, the histological performance of the networks was compared against the histological ground truth by expert evaluators for overall tissue appearance at a low magnification and for detailed histological accuracy at high magnifications, including morphological evaluation of the accuracy of overall tissue composition and structures, tissue type-specific performance, positioning of cells and nuclei, nuclear morphology, nuclear chromasia, and overall color representation.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2023.100725.

ACKNOWLEDGMENTS

Prof. Tapio Visakorpi is thanked for kindly providing tissue material. Tissue processing and imaging were carried out with the support of UEF Cell and Tissue Imaging Unit, University of Eastern Finland, Biocenter Kuopio, and Biocenter Finland. Eija Rahunen, Taina Vihavainen, and Janne Capra are thanked for advice and technical assistance, and Kiia Koivusalo and Siiri Sirviö for assistance with imaging. The authors gratefully acknowledge support from CSC - IT Center for Science, whose high-performance computing services were used in this study. This work was supported by the ERA PerMed 2019–2022 ABCAP project (P.R., L.L.), Academy of Finland (L.L. 317871, 334774, P.R. 334782, 341967, 335976), Sigrid Jusélius Foundation (L.L.), the Cancer Foundation Finland (L.L., P.R.), the Ida Montin Foundation (S.K. 20220044), University of Turku Graduate School (UTUGS) (U.K.), and North Savo Cancer Foundation (S.K. 20220020).

AUTHOR CONTRIBUTIONS

Conceptualization, P.R. and L.L.; Methodology, U.K., S.K., M.V., L.L., and P.R.; Software, U.K., and M.V.; Validation, U.K., S.K., M.V., and L.L.; Formal analysis, U.K.; Investigation, U.K., and S.K.; Resources, L.L., and P.R.; Data Curation, U.K., S.K., and P.R.; Writing – Original Draft, U.K. and S.K.; Writing – Review & Editing, U.K., S.K., M.V., L.L., and P.R.; Visualization, U.K., S.K., and M.V.; Supervision, L.L. and P.R.; Project Administration, L.L. and P.R.; Funding Acquisition, L.L. and P.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: October 27, 2022
Revised: December 23, 2022
Accepted: March 8, 2023
Published: April 7, 2023

REFERENCES

1. Titford, M. (2009). Progress in the development of microscopical techniques for diagnostic pathology. J. Histotechnol. *32*, 9–19. https://doi.org/10.1179/his.2009.32.1.9.
2. Chan, J.K.C. (2014). The wonderful colors of the hematoxylin-eosin stain in diagnostic surgical pathology. Int. J. Surg. Pathol. *22*, 12–32. https://doi.org/10.1177/1066896913517939.
3. Prezja, F., Pölönen, I., Äyrämö, S., Ruusuvuori, P., and Kuopio, T. (2022). H&E multi-laboratory staining variance exploration with machine learning. Appl. Sci. *12*, 7511. https://doi.org/10.3390/app12157511.
4. Pang, Y., Lin, J., Qin, T., and Chen, Z. (2022). Image-to-image translation: methods and applications. IEEE Trans. Multimedia *24*, 3859–3881. https://doi.org/10.1109/tmm.2021.3109419.
5. Jin, C.B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., and Cui, X. (2019). Deep CT to MR synthesis using paired and unpaired data. Sensors *19*, 2361. https://doi.org/10.3390/s19102361.
6. Brou Boni, K.N.D., Klein, J., Gulyban, A., Reynaert, N., and Pasquier, D. (2021). Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with 0-unpaired data. Med. Phys. *48*, 3003–3010. https://doi.org/10.1002/mp.14866.
7. Kawahara, D., and Nagata, Y. (2021). T1-weighted and T2-weighted MRI image synthesis with convolutional generative adversarial networks. Rep. Pract. Oncol. Radiother. *26*, 35–42. https://doi.org/10.5603/rpOr.a2021.0005.
8. MRCAT Brain. https://www.philips.fi/healthcare/product/HCNMRF320/mrcat-brain-mr-rt-clinical-application
9. Automatic segmentation service. https://www.mvision.ai/product/
10. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. *115*, 211–252. https://doi.org/10.1007/s11263-015-0816-y.
11. Wu, K., Chen, X., and Ding, M. (2014). Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. Optik *125*, 4057–4063. https://doi.org/10.1016/j.ijleo.2014.01.114.
12. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
13. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., Van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A.W.M., the CAMELYON16 Consortium, Hermsen, M., Manson, Q.F., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA *318*, 2199–2210. https://doi.org/10.1001/jama.2017.14585.
14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. Commun. ACM *63*, 139–144. https://doi.org/10.1145/3422622.
15. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. Adv. Neural Inf. Process. Syst. *29*.
16. Cohen, J.P., Luck, M., and Honari, S. (2018). Distribution matching losses can hallucinate features in medical image translation. In International Conference on Medical image computing and computer-assisted intervention, pp. 529–536. https://doi.org/10.1007/978-3-030-00928-1_60.
17. Zhu, J.Y., Park, T., Isola, P., and Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pp. 2223–2232. https://doi.org/10.1109/ICCV.2017.244.
18. de Bel, T., Bokhorst, J.M., van der Laak, J., and Litjens, G. (2021). Residual cyclegan for robust domain transformation of histopathological tissue slides. Med. Image Anal. *70*, 102004. https://doi.org/10.1016/j.media.2021.102004.
19. de Bel, T., Hermsen, M., Kers, J., van der Laak, J., and Litjens, G. (2018). Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology. In International Conference on Medical Imaging with Deep Learning.
20. Xu, Z., Moro, C.F., Bozóky, B., and Zhang, Q. (2019). GAN-based virtual re-staining: a promising solution for whole slide image analysis. Preprint at arXiv. https://doi.org/10.48550/arXiv.1901.04059.
21. Rivenson, Y., de Haan, K., Wallace, W.D., and Ozcan, A. (2020). Emerging advances to transform histopathology using virtual staining. BME Front. *2020*. https://doi.org/10.34133/2020/9647163.
22. Bayramoglu, N., Kaakinen, M., Eklund, L., and Heikkila, J. (2017). Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 64–71. https://doi.org/10.1109/ICCVW.2017.15.
23. Rivenson, Y., Liu, T., Wei, Z., Zhang, Y., de Haan, K., and Ozcan, A. (2019). PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. Light Sci. Appl. *8*, 23. https://doi.org/10.1038/s41377-019-0129-y.
24. Rana, A., Lowe, A., Lithgow, M., Horback, K., Janovitz, T., Da Silva, A., Tsai, H., Shanmugam, V., Bayat, A., and Shah, P. (2020). Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. JAMA Netw. Open *3*, e205111. https://doi.org/10.1001/jamanetworkopen.2020.5111.
25. Isola, P., Zhu, J.Y., Zhou, T., and Efros, A.A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125–1134. https://doi.org/10.1109/CVPR.2017.632.
26. Bayat, A., Anderson, C., and Shah, P. (2021, February). Automated end-to-end deep learning framework for classification and tumor localization from native non-stained pathology images. Medical Imaging 2021: Image Process. *11596*, 43–54. https://doi.org/10.1117/12.2582303.
27. Zhang, Y., de Haan, K., Rivenson, Y., Li, J., Delis, A., and Ozcan, A. (2020). Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. Light Sci. Appl. *9*, 78. https://doi.org/10.1038/s41377-020-0315-y.
28. Koivukoski, S., Khan, U., Ruusuvuori, P., and Latonen, L. (2023). Unstained tissue imaging and virtual hematoxylin and eosin staining of histological whole slide images. Lab. Invest. *103*, 100070. https://doi.org/10.1016/j.labinv.2023.100070.
29. Dong, R., Pan, X., and Li, F. (2019). DenseU-net-based semantic segmentation of small objects in urban remote sensing images. IEEE Access *7*, 65347–65356. https://doi.org/10.1109/ACCESS.2019.2917952.
30. Knoblaugh, S.E., Adissu, H.A., McKerlie, C., and Cardiff, R.D. (2021). Male reproductive system. Pathology of Genetically Engineered and Other Mutant Mice, 431–461.
31. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., and Rajpoot, N. (2019). Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. *58*, 101563. https://doi.org/10.1016/j.media.2019.101563.
32. Rogers, A.B., and Dintzis, R.Z. (2012). Liver and gallbladder. In Comparative Anatomy and Histology (Academic Press), pp. 193–201. https://doi.org/10.1016/B978-0-12-381361-9.00013-5.
33. Linden, M., Ward, J.M., and Cherian, S. (2012). Hematopoietic and lymphoid tissues. In Comparative Anatomy and Histology (Academic Press), pp. 309–338. https://doi.org/10.1016/B978-0-12-381361-9.00019-6.
34. Treuting, P.M., and Kowalewska, J. (2012). Urinary system. In Comparative Anatomy and Histology (Academic Press), pp. 229–251. https://doi.org/10.1016/B978-0-12-381361-9.00016-0.
35. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826. https://doi.org/10.1109/CVPR.2016.308.

36. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

37. Fischer, A.H., Jacobson, K.A., Rose, J., and Zeller, R. (2008). Hematoxylin and eosin staining of tissue and cell sections. Cold Spring Harb. Protoc. *2008*. pdb.prot4986.

38. Mayerich, D., Walsh, M.J., Kadjacsy-Balla, A., Ray, P.S., Hewitt, S.M., and Bhargava, R. (2015). Stain-less staining for computed histopathology. Technology *3*, 27–31. https://doi.org/10.1142/S2339547815200010.

39. Lahiani, A., Klaiman, E., and Grimm, O. (2018). Enabling histopathological annotations on immunofluorescent images through virtualization of hematoxylin and eosin. J. Pathol. Inform. *9*, 1. https://doi.org/10.4103/jpi.jpi_61_17.

40. Rana, A., Yauney, G., Lowe, A., and Shah, P. (2018). Computational histological staining and destaining of prostate core biopsy RGB images with generative adversarial neural networks. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 828–834. https://doi.org/10.1109/ICMLA.2018.00133.

41. Valkonen, M., Högnäs, G., Bova, G.S., and Ruusuvuori, P. (2021). Generalized fixation invariant nuclei detection through domain adaptation based deep learning. IEEE J. Biomed. Health Inform. *25*, 1747–1757. https://doi.org/10.1109/JBHI.2020.3039414.

42. Kim, J., Kim, M., Kang, H., and Lee, K. (2019). U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. Preprint at arXiv. https://doi.org/10.48550/arXiv.1907.10830.

43. Lin, J., Pang, Y., Xia, Y., Chen, Z., and Luo, J. (2020). Tuigan: learning versatile image-to-image translation with two unpaired images. In European Conference on Computer Vision, pp. 18–35. https://doi.org/10.1007/978-3-030-58548-8_2.

44. Lin, J., Wang, Y., Chen, Z., and He, T. (2020). Learning to transfer: unsupervised domain translation via meta-learning. Proc. AAAI Conf. Artif. Intell. *34*, 11507–11514. https://doi.org/10.1609/aaai.v34i07.6816.

45. Latonen, L., Scaravilli, M., Gillen, A., Hartikainen, S., Zhang, F.P., Ruusuvuori, P., Kujala, P., Poutanen, M., and Visakorpi, T. (2017). In vivo expression of mir-32 induces proliferation in prostate epithelium. Am. J. Pathol. *187*, 2546–2557. https://doi.org/10.1016/j.ajpath.2017.07.012.

46. Scaravilli, M., Koivukoski, S., Gillen, A., Bouazza, A., Ruusuvuori, P., Visakorpi, T., and Latonen, L. (2022). miR-32 promotes MYC-driven prostate cancer. Oncogenesis *11*, 11. https://doi.org/10.1038/s41389-022-00385-8.

47. Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2011). FSIM: a feature similarity index for image quality assessment. IEEE Trans. Image Process. *20*, 2378–2386. https://doi.org/10.1109/TIP.2011.2109730.

48. Hore, A., and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In 2010 20th international conference on pattern recognition. Istanbul: IEEE, pp. 2366–2369. https://doi.org/10.1109/ICPR.2010.579.