Title

Label-Free Virtual HER2 Immunohistochemical Staining of Breast Tissue using Deep Learning

Permalink

https://escholarship.org/uc/item/8105k50p

Authors

Bai, Bijie
Wang, Hongda
Li, Yuzhu
et al.

Publication Date

2022

DOI

10.34133/2022/9786242

Copyright Information

Peer reviewed

Research Article

Label-Free Virtual HER2 Immunohistochemical Staining of Breast Tissue using Deep Learning

Bijie Bai [ID],[1,2,3] Hongda Wang,[1,2,3] Yuzhu Li,[1,2,3] Kevin de Haan,[1,2,3] Francesco Colonnese,[4] Yujie Wan,[5] Jingyi Zuo,[4] Ngan B. Doan,[6] Xiaoran Zhang,[1] Yijie Zhang [ID],[1,2,3] Jingxi Li,[1,2,3] Xilin Yang,[1,2,3] Wenjie Dong,[7] Morgan Angus Darrow,[8] Elham Kamangar,[8] Han Sung Lee,[8] Yair Rivenson,[1,2,3] and Aydogan Ozcan [ID] [1,2,3,9]

[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA
[2]Bioengineering Department, University of California, Los Angeles 90095, USA
[3]California NanoSystems Institute (CNSI), University of California, Los Angeles, CA, USA
[4]Computer Science Department, University of California, Los Angeles, CA, USA
[5]Physics and Astronomy Department, University of California, Los Angeles, CA 90095, USA
[6]Translational Pathology Core Laboratory, University of California, Los Angeles, CA 90095, USA
[7]Statistics Department, University of California, Los Angeles, CA 90095, USA
[8]Department of Pathology and Laboratory Medicine, University of California at Davis, Sacramento, CA 95817, USA
[9]Department of Surgery, University of California, Los Angeles, CA 90095, USA

Correspondence should be addressed to Aydogan Ozcan; ozcan@ucla.edu

The immunohistochemical (IHC) staining of the human epidermal growth factor receptor 2 (HER2) biomarker is widely practiced in breast tissue analysis, preclinical studies, and diagnostic decisions, guiding cancer treatment and investigation of pathogenesis. HER2 staining demands laborious tissue treatment and chemical processing performed by a histotechnologist, which typically takes one day to prepare in a laboratory, increasing analysis time and associated costs. Here, we describe a deep learning-based virtual HER2 IHC staining method using a conditional generative adversarial network that is trained to rapidly transform autofluorescence microscopic images of unlabeled/label-free breast tissue sections into bright-field equivalent microscopic images, matching the standard HER2 IHC staining that is chemically performed on the same tissue sections. The efficacy of this virtual HER2 staining framework was demonstrated by quantitative analysis, in which three board-certified breast pathologists blindly graded the HER2 scores of virtually stained and immunohistochemically stained HER2 whole slide images (WSIs) to reveal that the HER2 scores determined by inspecting virtual IHC images are as accurate as their immunohistochemically stained counterparts. A second quantitative blinded study performed by the same diagnosticians further revealed that the virtually stained HER2 images exhibit a comparable staining quality in the level of nuclear detail, membrane clearness, and absence of staining artifacts with respect to their immunohistochemically stained counterparts. This virtual HER2 staining framework bypasses the costly, laborious, and time-consuming IHC staining procedures in laboratory and can be extended to other types of biomarkers to accelerate the IHC tissue staining used in life sciences and biomedical workflow.

1. Introduction

The immunohistochemical (IHC) staining of tissue sections plays a pivotal role in the evaluation process of a broad range of diseases. Since its first implementation in 1941 [1], a great variety of IHC biomarkers have been validated and employed in clinical and research laboratories for characterization of specific cellular events [2], e.g., the nuclear protein Ki-67 associated with cell proliferation [3], the cellular tumor antigen P53 associated with tumor formation [4], and the human epidermal growth factor receptor 2 (HER2) associated with aggressive breast tumor development [5].

Due to its capability of selectively identifying targeted biomarkers, IHC staining of tissue has been established as one of the gold standards for tissue analysis and diagnostic decisions, guiding disease treatment and investigation of pathogenesis [6–8].

Though widely used, the IHC staining of tissue still requires a dedicated laboratory infrastructure and skilled operators (histotechnologists) to perform laborious tissue preparation steps and is therefore time-consuming and costly. Recent years have seen rapid advances in deep learning-based virtual staining techniques, providing promising alternatives to the traditional histochemical staining workflow by computationally staining the microscopic images captured from label-free thin tissue sections, bypassing the laborious and costly chemical staining process. Such label-free virtual staining techniques have been demonstrated using autofluorescence imaging [9, 10], quantitative phase imaging [11], and light scattering imaging [12], among others [13–15], and have successfully created multiple types of histochemical stains, e.g., hematoxylin and eosin (H&E) [9–14], Masson's trichrome [9–11], and Jones silver stains [9–11]. These previous works did not perform any virtual IHC staining and mainly focused on the generation of structural tissue staining, which enhances the contrast of specific morphological features in tissue sections. In a related line of research, deep learning has also enabled the prediction of biomarker status (e.g., Ki-67 [16] and β-amyloid [17]) and tumor prognostic from H&E-stained microphotographs of various lesions including hepatocellular carcinoma [18], breast cancer [19–23], bladder cancer [24], thyroid cancer [25, 26], melanoma [27], and neuropathologic diseases [17]. These studies highlight a possible correlation between the presence of specific biomarkers and morphological microscopic changes in the tissue; however, they do not provide an alternative to IHC stained tissue images that reveal subcellular biomarker information for pathologists' diagnostic inspection for inter- and intracellular signatures such as cytoplasmic and nuclear details [28].

Here, we present a deep learning-based label-free virtual IHC staining method (Figure 1), which transforms autofluorescence microscopic images of unlabeled tissue sections into bright-field equivalent images, matching the standard IHC stained images of the same tissue samples. In this study, we specifically focused on the IHC staining of HER2, which is an important cell surface receptor protein that is involved in regulating cell growth and differentiation [29, 30]. Assessing the level of HER2 expression in breast tissue, i.e., HER2 status, is routinely practiced based on the HER2 IHC staining of the formalin-fixed, paraffin-embedded (FFPE) tissue sections and helps predict the prognosis of breast cancer and its response to HER2-directed immunotherapies [5, 30–34]. For example, the intracellular and extracellular studies of HER2 have led to the development of pharmacological anti-HER2 agents that benefit the treatment of HER2-positive tumors [35–39]. Further efforts are being made to develop new pharmacological solutions that can counter HER2-directed-drug resistance and improve treatment outcomes in clinical trials [40–43]. With numerous animal models established for preclinical studies and life sciences

related research, a deeper understanding of the oncogene, biological functionality, and drug resistance mechanisms of HER2 is being explored [44–48]. In addition to these, HER2 biomarker was also used as an essential tool in developing and testing of novel biomedical imaging [49, 50], statistics [51], and spatial transcriptomics [52] methods.

The presented virtual HER2 staining method is based on a deep learning-enabled image-to-image transformation, using a conditional generative adversarial network (GAN), as shown in Figure 2. Once the training phase was completed, two blinded quantitative studies were performed using new breast tissue sections with different HER2 scores to demonstrate the efficacy of our virtual HER2 staining framework. For this purpose, we used the semi-quantitative Dako HercepTest scoring system [53], which involves assessing the percentage of tumor cells that exhibit membranous staining for HER2 along with the intensity of the staining. The results are reported as 0 (negative), 1+ (negative), 2+ (weakly positive/equivocal), and 3+ (positive). In the first study, three board-certified breast pathologists blindly graded the HER2 scores of virtually stained HER2 whole slide images (WSIs) as well as their IHC stained standard counterparts. Our results and the statistical analysis revealed that determining the HER2 status based on our virtual HER2 WSIs is as accurate as standard analysis based on the chemically prepared IHC HER2 slides. In the second study, the same pathologists rated the staining quality of both virtual HER2 and standard IHC HER2 images using different metrics, i.e., nuclear detail, membrane clearness, background staining, and staining artifacts. This study revealed that at least two pathologists out of the three agreed that there is no statistically significant difference between the virtual HER2 staining image quality and the standard IHC HER2 staining image quality in the level of nuclear detail, membrane clearness, and absence of staining artifacts. Additional feature-based quantitative assessments also confirmed the high degree of agreement between the virtually generated HER2 images and their standard IHC-stained counterparts, in terms of both nucleus and membrane stain features.

The presented framework achieved the first demonstration of label-free virtual IHC staining and bypasses the costly, laborious, and time-consuming IHC staining procedures that involve toxic chemical compounds. This virtual HER2 staining technique has the potential to be extended to virtual staining of other biomarkers and may accelerate the IHC-based tissue analysis workflow in life sciences and biomedical applications, while also enhancing the repeatability and standardization of IHC staining.

2. Results

2.1. Label-Free Virtual HER2 Staining of Breast Tissue. We demonstrated our virtual HER2 staining method by training deep neural network (DNN) models with a dataset of 25 breast tissue sections collected from 19 unique patients, constituting in total 20,910 image patches, each with 1024×1024 pixels. Once a DNN model was trained, it virtually stained the unlabeled tissue sections using their autofluorescence microscopic images captured with DAPI, FITC,

Unstained breast tissue section — HER2 staining — Immunohistochemically stained HER2

DAPI FITC TxRed CY5

Autofluorescence images — Deep neural network — Virtually stained HER2

1 mm

200 μm

50 μm

(a)

Autofluorescence image — Trained network model — Virtually stained HER2 — Immunohistochemically stained HER2
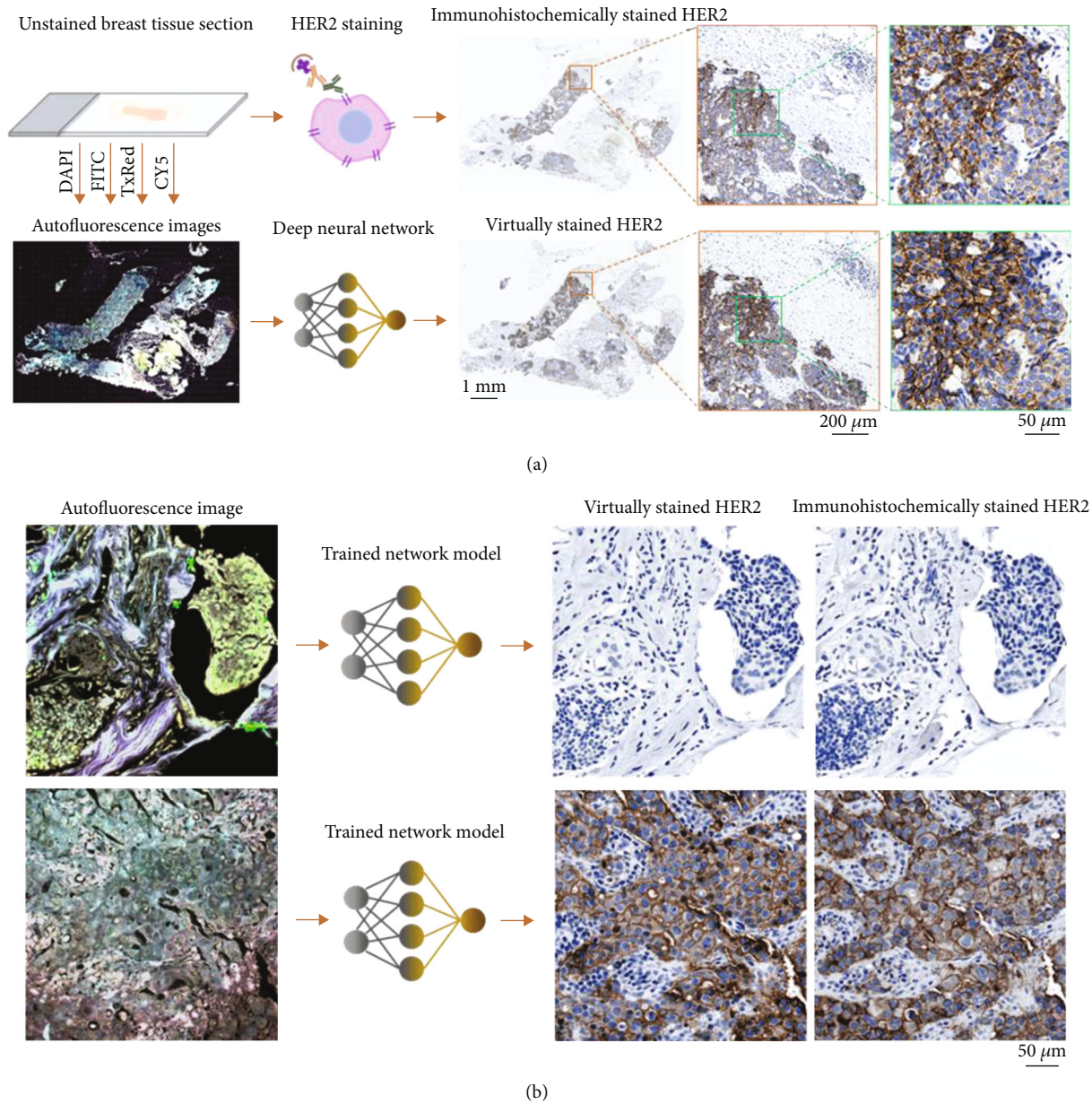
Trained network model

50 μm

(b)

FIGURE 1: Virtual HER2 staining of unlabeled tissue sections via deep learning. (a) The standard immunohistochemical (IHC) HER2 staining (top) relies on tedious and costly tissue processing performed by histotechnologists, which typically takes ~1 day. A pretrained deep neural network enables virtual HER2 staining of unlabeled tissue sections (bottom). (b) Virtual HER2 staining transforms autofluorescence images of unlabeled tissues sections into bright-field equivalent images that match the images of standard IHC HER2 staining.

TxRed, and Cy5 filter cubes (see Methods section), matching the corresponding bright-field images of the same field-of-views, captured after standard IHC HER2 staining. In the network training and evaluation process, we employed a cross-validation approach. Separate network models were trained with different dataset divisions to generate 12 virtual HER2 WSIs for blind testing, i.e., 3 WSIs at each of the 4 HER2 scores (0, 1+, 2+, and 3+). Each virtual HER2 WSI corresponds to a unique patient that was not used during the network training phase. Note that all the tissue sections were obtained from existing tissue blocks, where the HER2 reference (ground truth) scores were provided by UCLA

Translational Pathology Core Laboratory (TPCL) under UCLA IRB 18-001029.

Figure 3 summarizes the comparison of the virtual HER2 images inferred by our DNN models against their corresponding IHC HER2 images captured from the same tissue sections after standard IHC staining. Both the WSIs and the zoomed-in regions show a high degree of agreement between virtual staining and standard IHC staining. These results indicate that a well-trained virtual staining network can reliably transform the autofluorescence images of unlabeled breast tissue sections into the bright-field equivalent, virtual HER2 images, which match their IHC HER2 stained
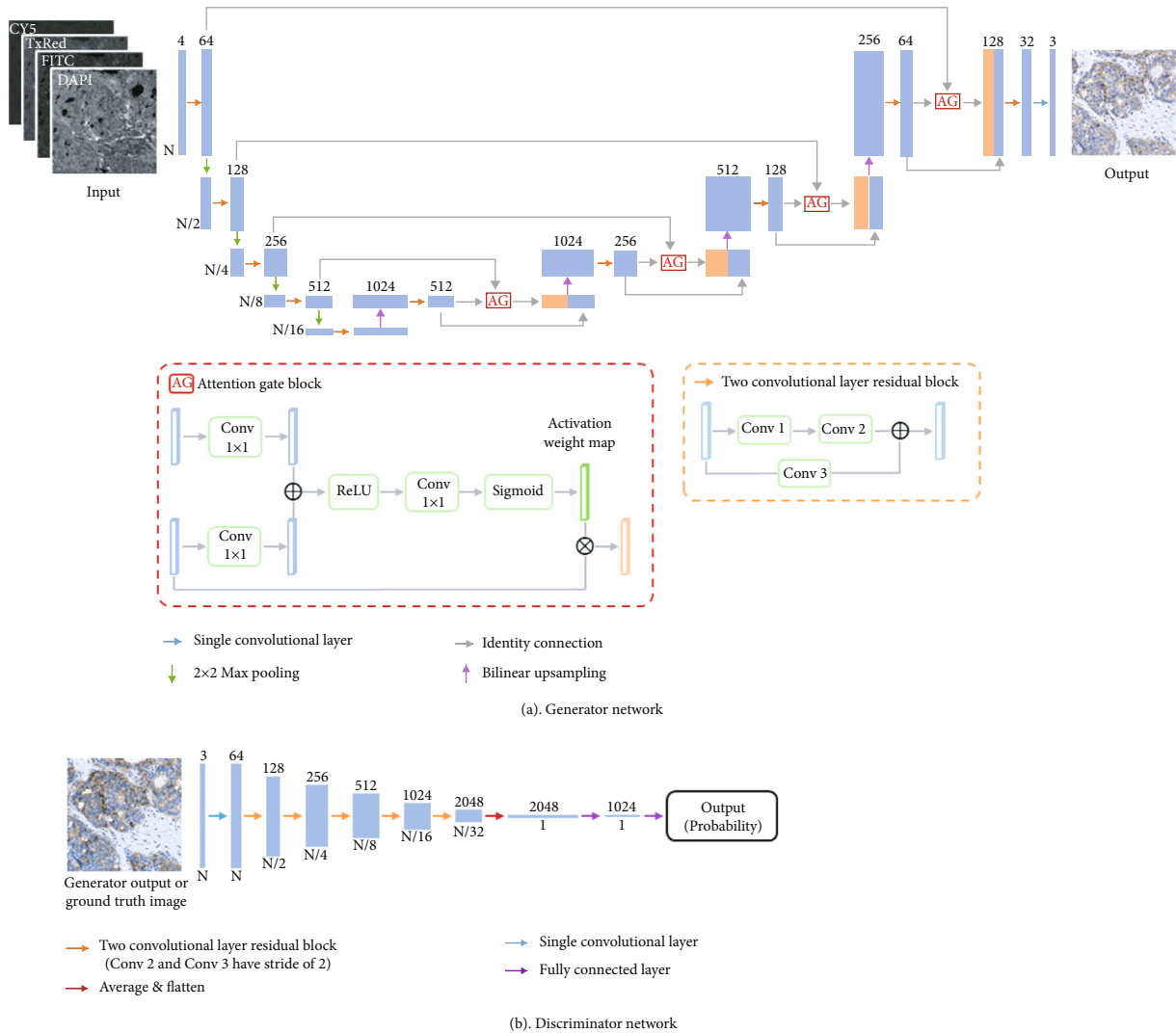
FIGURE 2: Virtual HER2 staining network. A GAN framework which consists of a generator model and a discriminator model was used to train the virtual HER2 staining network. (a) The generator uses an attention-gated U-net structure to map the label-free autofluorescence images into bright-field equivalent HER2 images. (b) The discriminator is a CNN composed of five successive two-convolutional-layer residual blocks and two fully connected layers (see Methods). Once the network models converge, only the generator model is used to infer the virtual HER2 images, which takes ~12 seconds for 1 mm^2 of tissue area.

counterparts, across all the HER2 statuses, 0, 1+, 2+, and 3+. Upon close examination, our board-certified pathologists confirmed that the comparison between the IHC and virtual HER2 images showed equivalent staining with no significant perceptible differences in intracellular features such as membrane clarity or nuclear details. In particular, the virtual staining network clearly produced the expected intensity and distribution of membranous HER2 staining (or lack thereof) in tumor cells. In HER2 positive (3+, Figures 3(a)–3(e)) breast cancers, both virtually stained and IHC stained images showed strong complete membranous staining in >10% of tumor cells, as well as dim cytoplasmic staining in tumor cells. None of the stromal and inflammatory cells showed false-positive staining, and the nuclear details of the tumor cells were comparable in both panels. In equivocal (2+, Figures 3(f)–3(j)) tumors, virtual images showed weak to moderate membranous staining in >10% of tumor

cells, providing the same amount of membranous staining of tumor cells in corresponding areas. HER2 negative (1+, Figures 3(k)–3(o)) tumors showed faint membranous staining in 10% or more of tumor cells. None of the stromal and inflammatory cells showed faint staining. HER2 negative (0, Figures 3(p)–3(t)) tumor showed no staining in the tumor cells.

2.2. Blind Evaluation and Quantification of Virtual HER2 Staining. Next, we evaluated the efficacy of the presented virtual HER2 staining framework with a quantitative blinded study in which the 12 virtual HER2 WSIs and their corresponding standard IHC HER2 WSIs were mixed and presented to three board-certified breast pathologists who graded the HER2 score (i.e., 3+, 2+, 1+, or 0) for each WSI without knowing if the image was from a virtual stain or standard IHC stain. Random image shuffling, rotation, and
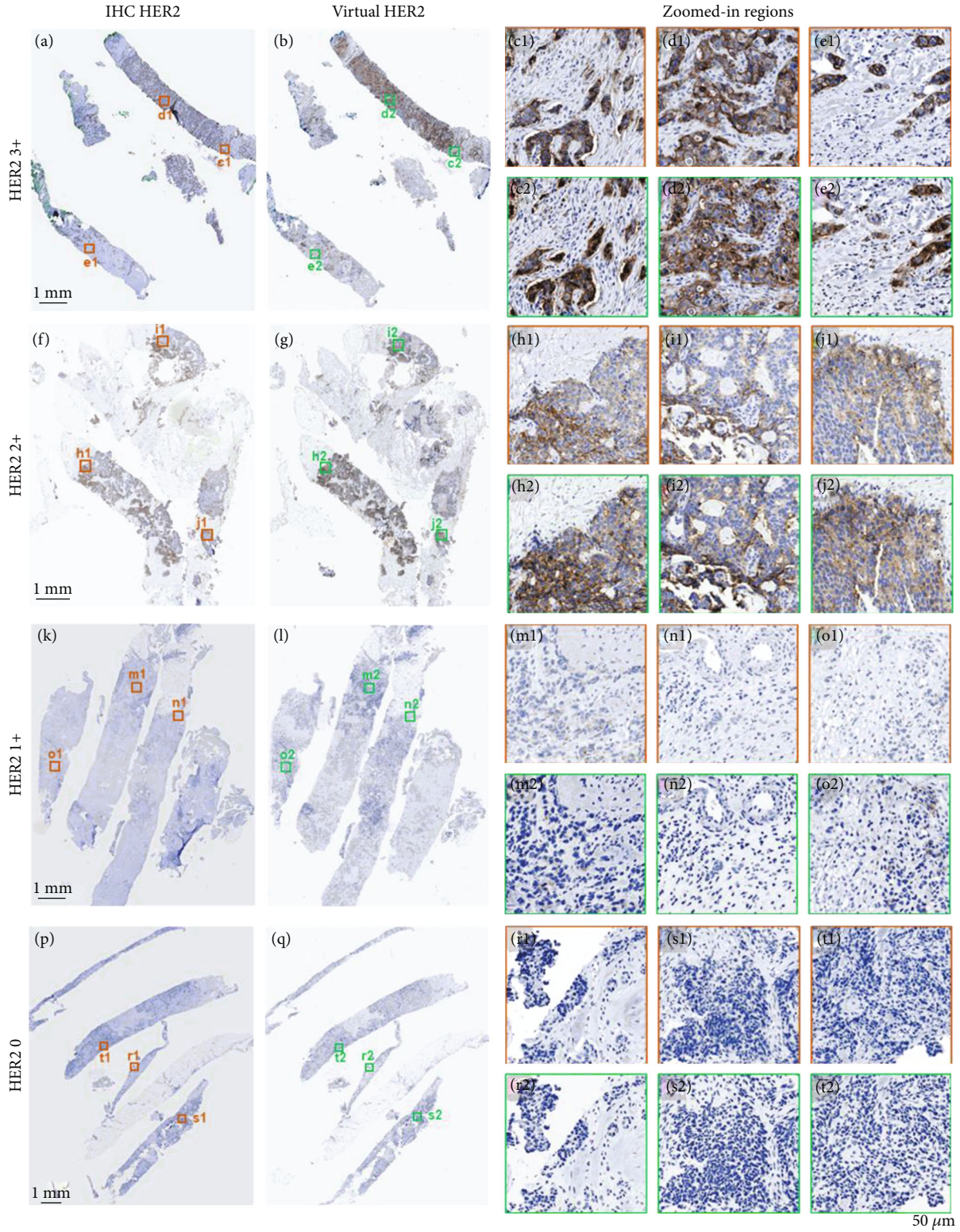
FIGURE 3: Comparison of virtual and standard IHC HER2 staining of breast tissue sections at different HER2 scores. (a), (f), (k), (p) Bright-field WSIs of standard IHC HER2 stained samples at a HER2 3+, f HER2 2+, k HER2 1+, and p HER2 0. (b), (g), (l), (q) Bright-field WSIs generated by virtual staining, corresponding to the same samples as (a), (f), (l), and (p), respectively. (c1)-(e1), (c2)-(e2) Zoomed-in regions of interest from (a) and (b) at a HER2 score of 3+. (h1)-(j1), (h2)-(j2) Zoomed-in regions of interest from (f) and (g) at a HER2 score of 2+. (m1)-(o1), (m2)-(o2) Zoomed-in regions of interest from (k) and (l) at a HER2 score of 1+. (r1)-(t1), (h2)-(t2) Zoomed-in regions of interest from (p) and (q) at a HER2 score of 0.

flipping were applied to the WSIs to promote blindness in evaluations. The HER2 scores of the virtual and the standard IHC WSIs that were blindly graded by the three pathologists are summarized in Figure 4 and compared to their reference, ground truth scores provided by UCLA TPCL. The confusion matrices of virtual HER2 WSIs (Figure 4(a)) and IHC HER2 WSIs (Figure 4(b)), each corresponding to $N = 36$ evaluations, reveal that our virtual HER2 staining approach achieved a similar level of accuracy for HER2 status assessment as the standard IHC staining. Close examination of these confusion matrices reveals that the sum of the diagonal elements of the virtual HER2-based evaluations (22) is higher than that of the IHC HER2 (19), showing that more cases were correctly scored based on virtual HER2 WSIs compared to those based on standard IHC HER2 WSIs. Furthermore, the sum of the absolute off-diagonal errors of virtual HER2-based evaluations (14) is smaller than that of the standard IHC HER2 (18). Based on the same confusion matrices shown in Figure 4, a Chi-square test was performed to compare the degree of agreement between virtual staining and standard IHC staining methods in HER2 scoring. The test results indicate that there is no statistically significant difference between the two methods ($P = 0.4752$, see Supplementary Table 1).

In addition to evaluating the efficacy of virtual staining in HER2 scoring, we also quantitatively evaluated the staining quality of the virtual HER2 images and compared them to the standard IHC HER2 images. In this blinded study, we randomly extracted 10 regions-of-interest (ROIs) from each of the 12 virtual HER2 WSIs and 10 ROIs at the same locations from each of their corresponding IHC HER2 WSIs, building a test set of 240 image patches. Each image patch has 8000×8000 pixels ($1.3 \times 1.3 \, \mathrm{mm}^2$), which was also randomly shuffled, rotated, and flipped before being reviewed by the same three pathologists. These pathologists were asked to grade the image quality of each ROI based on four predesignated feature metrics for HER2 staining: membrane clearness, nuclear detail, absence of excessive background staining, and absence of staining artifacts (Figure 5). The grade scale for each metric is from 1 to 4, with 4 representing perfect, 3 representing very good, 2 representing acceptable, and 1 representing unacceptable. Figure 5(a) summarizes the staining quality scores of virtual HER2 and standard IHC HER2 images based on our predefined feature metrics, which were averaged over all image patches and pathologists. Figures 5(b)–5(e) further compare the average quality scores at each of the 4 HER2 statuses under each feature metric. In Figure 5(b), the membrane clearness scores of HER2 negative ROIs are noted as "not applicable" since there is no staining of the cell membrane in HER2 negative samples. It is important to emphasize that the standard IHC HER2 images had an advantage in these comparisons because they were preselected: A significant percentage of the standard IHC HER2 tissue slides suffered from unacceptable staining quality issues (see Discussion and Supplementary Figure 1), and therefore they were excluded from our comparative studies in the first place. Nevertheless, the quality scores of virtual and standard IHC HER2 staining are very close to each other and fall within their standard

deviations (dashed lines in Figure 5). We also performed one-sided t-tests on each feature metric evaluated by board-certified pathologists to determine whether standard IHC HER2 images are statistically significantly better than the virtual HER2 images in staining quality. The t-test results showed that only for the metric of "absence of excessive background staining," two of the three pathologists reported a statistically significant improvement in the quality of the standard IHC staining compared to the virtual staining. For the rest of the feature metrics (i.e., nuclear details, membrane clearness, and staining artifacts), at least two of the three pathologists reported that the staining quality of the IHC HER2 images is not statistically significantly better than their virtual HER2 counterparts (Supplementary Table 2). Also note that the virtually stained HER2 images did not mislead the diagnosis at the whole slide level as also analyzed using the confusion matrices shown in Figure 4 and the Chi-square test reported in Supplementary Table 1.

Besides rating the staining quality of each ROI, the pathologists also graded a HER2 score for each ROI, the results of which are reported in Supplementary Figure 2. Each histogram in Supplementary Figure 2a summarizes the HER2 scores of the 10 ROIs extracted from each WSI evaluated by 3 pathologists (i.e., $N = 30$ evaluations). The reference (ground truth) HER2 scores of the corresponding WSIs are plotted as gray dashed lines. This analysis reveals that, for the majority of the patients, there is no discrepancy between HER2 scores evaluated from virtually generated ROIs and standard IHC stained ROIs. For the cases where there is a disagreement (e.g., Patients #5 and #11), the histograms of the virtual HER2 scores were centered closer to the reference HER2 scores (dashed lines) compared to the histograms of the standard IHC-based HER2 scores. It is important to also note that grading the HER2 scores from subsampled ROIs vs. from the WSI can yield different results due to the inhomogeneous nature of the tissue sections.

2.3. Feature-Based Quantitative Assessment of Virtual HER2 Staining. In addition to the pathologists' blind assessments of the virtual staining efficacy and the image quality, we further carried out a feature-based quantitative analysis of the virtually generated HER2 images compared to their IHC-stained counterparts. In this analysis, 8194 unique test image patches (each with a size of 1024×1024 pixels) were blindly selected for virtual staining. Due to the different staining features of each different HER2 status, these blind testing images were divided into two subsets for quantitative evaluation: one subset containing the images from HER2 0 and HER2 1+, $N = 4142$, and the other containing the images from HER2 2+ and HER2 3+, $N = 4052$. For each virtually stained HER2 image and its corresponding IHC HER2 image (ground truth), four feature-based quantitative evaluation metrics (specifically designed for HER2) were calculated based on the segmentation of nucleus stain and membrane stain (see the Methods section). These four feature-based evaluation metrics included the number of nuclei and the average nucleus area (in number of pixels) for quantifying the nucleus stain in each image as well as

Virtual HER2 staining confusion matrix

		Reference HER2 score			
		3+	2+	1+	0
Evaluated HER2 score (Virtual staining)	3+	9	5	0	0
	2+	0	4	3	0
	1+	0	0	6	6
	0	0	0	0	3

Sum of diagonal element = 22 (max = 36)
Sum of off-diagonal errors = 14

(a)

IHC HER2 staining confusion matrix

		Reference HER2 score			
		3+	2+	1+	0
Evaluated HER2 score (Chemical staining)	3+	3	2	0	0
	2+	6	7	0	1
	1+	0	0	3	2
	0	0	0	6	6

Sum of diagonal element = 19 (max = 36)
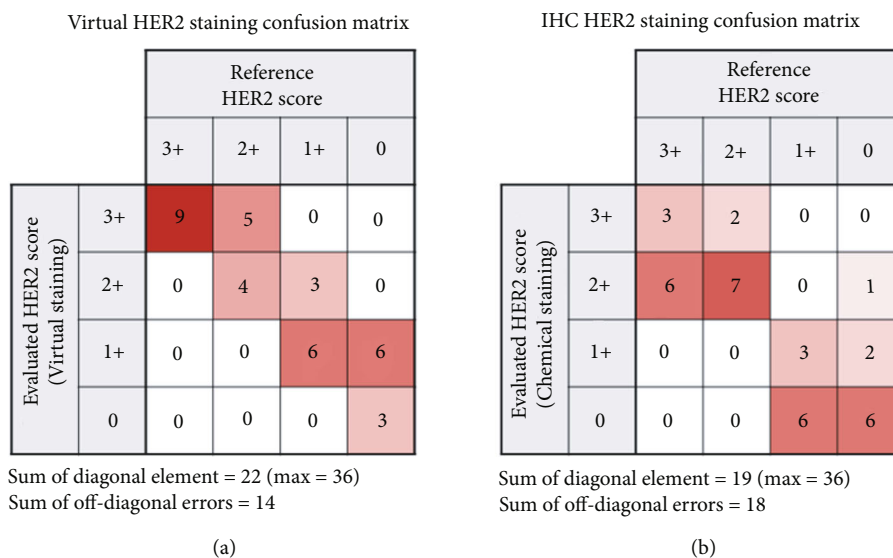Sum of off-diagonal errors = 18

(b)

FIGURE 4: Confusion matrices of HER2 scores. Each element in the matrices represents the number of WSIs with their HER2 scores evaluated by board-certified pathologists (rows) based on: (a) virtual HER2 staining or (b) standard IHC HER2 staining, compared to the reference (ground truth) HER2 scores provided by UCLA TPCL (columns).

the area under the characteristic curve and the membrane region connectedness [54, 55] for quantifying the membrane stain in each image (refer to the Methods section for details).

These feature-based quantitative evaluation results for the virtual HER2 images compared against their standard IHC counterparts are shown in Figure 6. This analysis demonstrated that the virtual HER2 staining feature metrics exhibit similar distributions and closely matching average values (dashed lines) compared to their standard IHC counterparts, in terms of both the nucleus and the membrane stains. By comparing the evaluation results of the HER2 positive group (2+ and 3+) against the HER2 negative group (0 and 1+), we observe similar distributions of nucleus features (i.e., the number of nuclei and average nucleus area) and higher levels of membrane stain, which correlates well with the higher HER2 scores as expected.

3. Discussion

We demonstrated a deep learning-enabled label-free virtual IHC staining method. By training a DNN model, our method generated virtual HER2 images from the autofluorescence images of unlabeled tissue sections, matching the bright-field images captured after standard IHC staining. Compared to chemically performing the IHC staining, our virtual HER2 staining method is rapid and simple to operate. The conventional IHC HER2 staining involves laborious sample treatment steps demanding a histotechnologist's periodic monitoring (see Supplementary Note 1), and this whole process typically takes one day before the slides can be reviewed by diagnosticians. In contrast, the presented virtual HER2 staining method bypasses these laborious and costly steps and generates the bright-field equivalent HER2 images computationally using the autofluorescence images captured from label-free tissue sections. After the training is complete (which is a one-time effort), the entire inference process using a virtual staining network only takes ~12 seconds for 1 mm^2 of tissue using a consumer-grade computer, which can be further improved by using faster hardware acceleration units.

Another advantage of the presented method is its capability of generating highly consistent and repeatable staining results, minimizing the staining variations that are commonly observed in standard IHC staining. The IHC HER2 staining procedure is delicate and laborious as it requires accurate control of time, temperature, and concentrations of the reagents at each tissue treatment step; in fact, it often fails to generate satisfactory stains. In our study, ~30% of the sample slides were discarded because of unsuccessful standard IHC staining and/or severe tissue damage even though the IHC staining was performed by accredited pathology labs. Supplementary Figure 1 shows two examples of the standard IHC staining failures we experienced, including complete tissue damage and false negative staining that failed to reflect the correct HER2 score. In contrast, our computational virtual staining approach does not rely on the chemical processing of the tissue and generates reproducible results, which is important for the standardization of the HER2 interpretation by eliminating commonly experienced staining variations and artifacts.

Since the autofluorescence input images of tissue slices were captured with standard filter sets installed on a conventional fluorescence microscope, the presented approach is ready to be implemented on existing fluorescence microscopes without hardware modifications or customized optical components. Our results showed that the combination of the four commonly used fluorescence filters (DAPI, FITC, TxRed, and Cy5) provided a very good baseline for the virtual HER2 staining performance. As an ablation study, we also quantitatively compared virtual staining networks that are trained with different autofluorescence input channels by calculating peak signal-to-noise ratio (PSNR) and
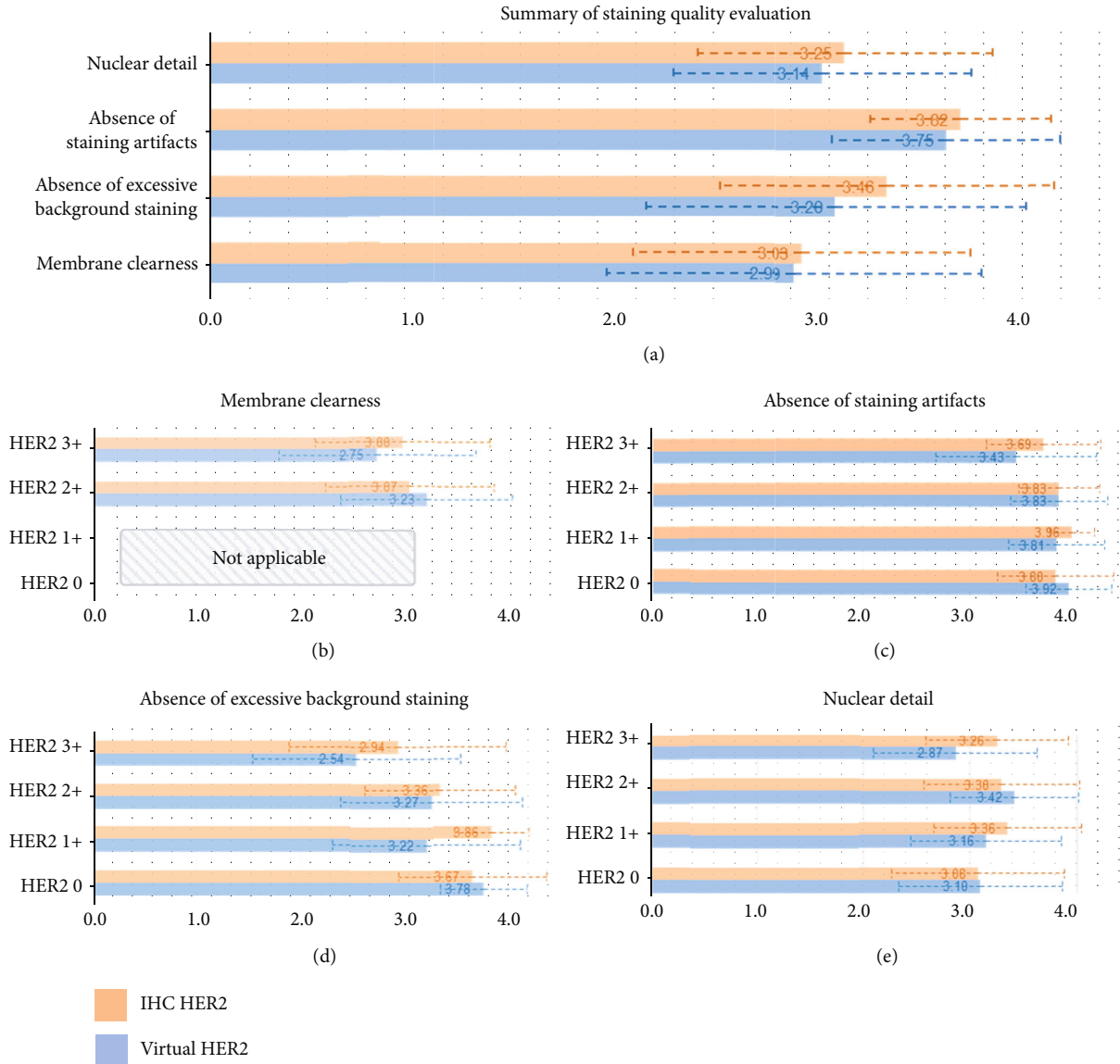
Figure 5: Comparisons of image quality of virtual HER2 and standard IHC HER2 staining. (a) Quality scores of virtual HER2 and standard IHC HER2 images calculated based on 4 different feature metrics: nuclear details, absence of staining artifacts, absence of excessive background staining, and membrane clearness. Each value was averaged over all the image patches and pathologists. (b)–(e) Detailed comparisons of quality scores under each feature metric at different HER2 scores. The grade scale applied for each metric is 1 to 4: 4 for perfect, 3 for very good, 2 for acceptable, and 1 for unacceptable. The standard deviations are plotted by dashed lines.

structural similarity index (SSIM) [56] between the network output and ground truth images (see Supplementary Figure 3). Since the staining of the cell membrane is an important assessment factor in HER2 status evaluation, we also performed color deconvolution [57] to split out the membrane stain channel (i.e., diaminobenzidine (DAB) stain) followed by calculating and comparing the SSIM scores (Supplementary Figure 4). These analyses revealed that the performance of the virtual staining network partially degraded with decreasing number of input autofluorescence channels, motivating the use of DAPI, FITC, TxRed, and Cy5 altogether (Supplementary Figure 3b).

The advantages of using the attention-gated GAN structure for virtual HER2 staining are illustrated by an addi-

tional comparative study, in which we trained and blindly tested four different network architectures including (1) the attention-gated GAN structure used in this work, (2) the same structure as ours with the residual connections removed, (3) the same structure as ours with the attention-gated blocks removed, and (4) an unsupervised cycleGAN [58, 59] framework. The training/validation/testing datasets and the training epochs were kept the same for all the four networks. After their training, we quantitatively compared these networks by calculating the PSNR, SSIM, and SSIM of the membrane stain ($SSIM_{DAB}$) between the network output and the ground truth images (see Supplementary Figure 5). Both the visual and numerical comparisons revealed that the attention-gated GAN used in this work is

N = 4142 images (Each with 1024×1024 pixels)

(a)

N = 4052 images (Each with 1024×1024 pixels)
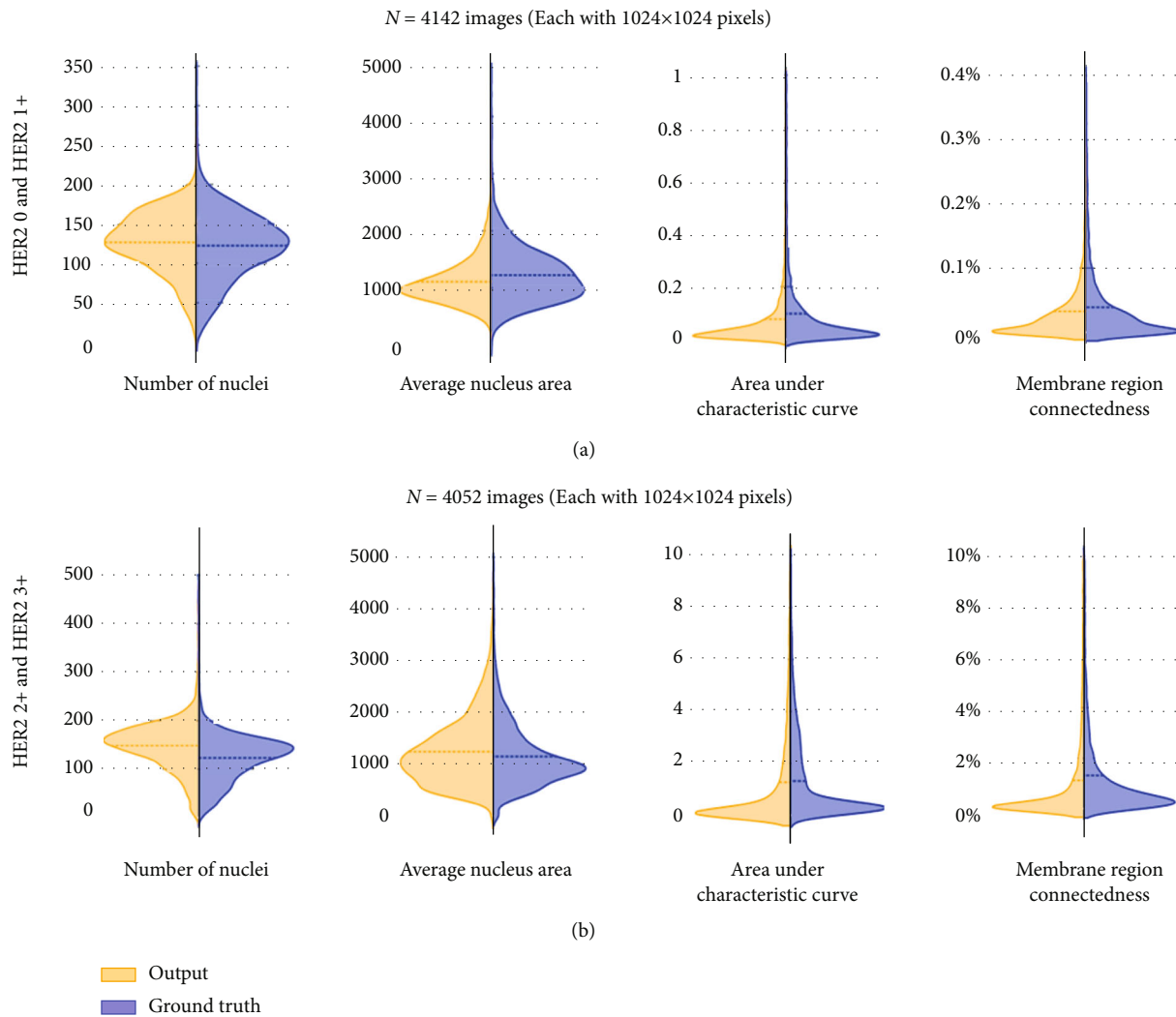

(b)

Output
Ground truth

FIGURE 6: Feature-based quantitative assessment of virtually stained HER2 images and standard IHC HER2 images. (a) Virtual HER2 features and standard IHC HER2 features, quantitatively compared for HER2 negative cases ($N = 4,142$ images) based on four different metrics. (b) Virtual HER2 features and standard IHC HER2 features, quantitatively compared for HER2 positive cases ($N = 4,052$ images) based on the same metrics.

the only network architecture that could provide consistently superior and accurate virtual staining results at various HER2 expression levels, while the other network architectures made some catastrophic staining errors in one or more testing FOVs, making them unacceptable for consistent inference across all HER2 statuses. In Supplementary Figures 6 and 7, we further compared the color distributions (see the Methods section) of the output images generated by these different network architectures against the corresponding ground truth images, including FOVs with strong HER2 expression (Supplementary Figure 6) and FOVs with weak HER2 expression (Supplementary Figure 7). These additional comparisons showed that the color histograms of the output images generated by our framework match with the standard IHC ground truth much more closely for both the membrane and nucleus stain channels, which again illustrates the advantages of using the attention-gated GAN architecture reported in this work.

The success of our virtual HER2 staining method relies on the processing of the complex spatial-spectral information that is encoded in the autofluorescence images of label-free tissue using convolutional neural networks. The presented virtual staining method can potentially be expanded to a wide range of other IHC stains. Though our virtual HER2 staining framework was demonstrated based on autofluorescence imaging of unlabeled tissue sections, other label-free microscopy modalities may also be utilized for this task, such as holography [11], fluorescence lifetime imaging [60, 61], and Raman microscopy [62]. In addition to generalizing to other types of IHC stains in the assessment of various biomarkers, this method can be further adapted to nonfixed fresh tissue samples or frozen sections, which can potentially provide real-time virtual IHC images for intraoperative consultation during surgical operations.

To the best of our knowledge, our results (placed in arXiv [63] on December 8, 2021) constitute the first

demonstration of label-free virtual IHC staining, and we believe that this framework opens up new avenues for various applications in life sciences and biomedical diagnostics and can potentially transform the traditional IHC staining workflow.

4. Methods

4.1. Sample Preparation and Standard IHC Staining. The unlabeled breast tissue blocks were provided by the UCLA TPCL under UCLA IRB 18-001029 and were cut into $4 \mu m$ thin sections. The FFPE thin sections were then deparaffinized and covered with glass coverslips. After acquiring the autofluorescence microscopic images, the unlabeled tissue sections were sent to accredited pathology labs for standard IHC HER2 staining, which was performed by UCLA TPCL and the Department of Anatomic Pathology of Cedars-Sinai Medical Center in Los Angeles, USA. The IHC HER2 staining protocol provided by UCLA TPCL is described in Supplementary Note 1.

4.2. Image Data Acquisition. The autofluorescence images of the unlabeled tissue sections were captured using a standard fluorescence microscope (IX-83, Olympus) with a $\times 40$/ 0.95NA (UPLSAPO, Olympus) objective lens. Four fluorescent filter cubes, including DAPI (Semrock DAPI-5060C-OFX, EX 377/50 nm, EM 447/60 nm), FITC (Semrock FITC-2024B-OFX, EX 485/20 nm, EM 522/24 nm), TxRed (Semrock TXRED-4040C-OFX, EX 562/40 nm, EM 624/ 40 nm), and Cy5 (Semrock CY5-4040C-OFX, EX 628/ 40 nm, EM 692/40 nm) were used to capture the autofluorescence images at different excitation-emission wavelengths. Each autofluorescence image was captured with a scientific complementary metal-oxide-semiconductor (sCMOS) image sensor (ORCA-flash4.0 V2, Hamamatsu Photonics) with an exposure time of 150 ms, 500 ms, 500 ms, and 1000 ms for DAPI, FITC, TxRed, and Cy5 filters, respectively. The image acquisition process was controlled by μManager (version 1.4) microscope automation software [64]. After the standard IHC HER2 staining is complete, the bright-field WSIs were acquired using a slide scanner microscope (AxioScan Z1, Zeiss) with a $\times 20$/0.8NA objective lens (Plan-Apo).

4.3. Image Preprocessing and Registration. The matching of the autofluorescence (network input) and the bright-field IHC HER2 (network ground truth) image pairs is critical for the successful training of an image-to-image transformation network. The image processing workflow for preparing the training dataset for our virtual HER2 staining network is described in Supplementary Figure 8, which was implemented in MATLAB (MathWorks). First, the autofluorescence images (before the IHC staining) and the whole-slide bright-field images (after the IHC staining) of the same tissue sections were stitched into WSIs (Supplementary Figure 8(a)) and globally co-registered by detecting and matching the speeded up robust features (SURF) points [65] (Supplementary Figure 8(b)). Then, these coarsely matched autofluorescence and bright-field WSIs were cropped into pairs of image tiles of 1024 ×

1024 pixels (Supplementary Figure 8(c)). These image pairs were not accurately matched at the pixel level due to optical aberrations and morphological changes of the tissue structure during the standard (laborious) IHC staining procedures. In order to calculate the transformation between the autofluorescence image and its bright-field counterpart using a correlation-based elastic registration algorithm [66], a registration model [9] needs to be trained to match the style of the autofluorescence images to the style of the bright-field images (Supplementary Figure 8d). This registration network used the same architecture as our virtual staining network. Following the image style transformation using the registration network (Supplementary Figure 8(e)), the pyramid elastic image registration algorithm [66, 67] was performed to hierarchically match the local features of the sub-image blocks and calculate the transformation maps. The transformation maps were then applied to correct for the local wrappings of the ground truth images (Supplementary Figure 8(f)) which were then better matched to their autofluorescence counterparts. This training-registration process (Supplementary Figures 8(d)-8(f)) was repeated 3-5 times until the autofluorescence input and the bright-field ground truth image patches were accurately matched at the single pixel-level (Supplementary Figure 8(g)). At last, a manual data cleaning process was performed to remove image pairs with artifacts such as tissue-tearing (during the standard chemical staining process) or defocusing (during the imaging process).

4.4. Virtual HER2 Staining Network Architecture and Training Schedule. In this work, a GAN-based network model [68] was employed to perform the transformation from the 4-channel label-free autofluorescence images (DAPI, FITC, TxRed, and Cy5) to the corresponding bright-field virtual HER2 images, as shown in Figure 2. This GAN framework includes (1) a generator network that creates virtually stained HER2 images by learning the statistical transformation between the input autofluorescence images and the corresponding bright-field IHC stained HER2 images (ground truth) and (2) a discriminator network that learns to discriminate the virtual HER2 images created by the generator from the actual IHC stained HER2 images. The generator and the discriminator were alternatively optimized and simultaneously improved through this competitive training process. Specifically, the generator (G) and discriminator (D) networks were optimized to minimize the following loss functions:

$$
\begin{aligned}
l_{\text{generator}} = {} & \alpha \times L_1 \{ I_{\text{target}}, G(I_{\text{input}}) \} - \lambda \\
& \times \log \left((1 + SSIM \{ I_{\text{target}}, G(I_{\text{input}}) \})/2 \right) \\
& + \gamma \times BCE \{ D(G(I_{\text{input}})), 1 \},
\end{aligned}
$$

$$
l_{\text{discriminator}} = BCE \{ D(G(I_{\text{input}})), 0 \} + BCE \{ D(I_{\text{target}}), 1 \},
$$

$$(1)$$

where $G(\bullet)$ represents the generator inference; $D(\bullet)$ represents the probability of being a real, actually-stained IHC

image predicted by the discriminator; I_{input} denotes the input label-free autofluorescence images; and I_{target} denotes the ground truth, standard IHC stained image. The coefficients $(\alpha, \lambda, \gamma)$ in $l_{\text{generator}}$ were empirically set as (10, 0.2, and 0.5) to balance the pixel-wise smooth L_1 error [69] of the generator output with respect to its ground truth, SSIM loss [56] of the generator output, and the binary cross-entropy (BCE) loss of the discriminator predictions of the output image. Compared to using the mean squared error (MSE) loss, the smooth L_1 loss is a robust estimator that prevents exploding gradients by using MSE around zero and mean absolute error (MAE) in other parts [70]. Specifically, smooth L_1 loss between two images A and B is defined as

$$
L_1\{A, B\} = \frac{1}{M \times N} \left(\sum_{\substack{m,n \\ |A(m,n)-B(m,n)|<\beta}} 0.5 \times \frac{(A(m, n) - B(m, n))^2}{\beta} + \sum_{\substack{m,n \\ |A(m,n)-B(m,n)|\geq\beta}} |A(m, n) - B(m, n)| - 0.5\beta \right),
\tag{2}
$$

where m and n are the pixel indices, the $M \times N$ represents the total number of pixels in each image, and β was set to 1 in our case.

The SSIM of two images is defined as [56]

$$
SSIM\{A, B\} = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)},
\tag{3}
$$

where μ_A and μ_B are the mean values of the images A and B, σ_A^2 and σ_B^2 are the variance of images A and B, and σ_{AB} is the covariance between images A and B. c_1 and c_2 were set to be 0.01^2 and 0.03^2, respectively [56].

The BCE with logits loss used in our network is defined as

$$
BCE\{p, q\} = -[q \times \log(\text{sigmoid}(p)) + (1 - q) \times \log(1 - \text{sigmoid}(p))],
\tag{4}
$$

where p represents the discriminator predictions and q represents the actual labels (0 or 1).

As shown in Figure 2(a), the generator network was built following the attention U-Net architecture [71] with 4 resolution levels, which can map the label-free autofluorescence images into the HER2 stained images by learning the transformations of spatial features at different spatial scales, catching both the high-resolution local features at shallower levels and the larger scale global context at deeper levels. Our attention U-Net structure is composed of a downsampling path and an upsampling path that are symmetric to each other. The downsampling path contains four downsampling convolutional blocks, each consisting of a two-convolutional-layer residual block, followed by a leaky rectified linear unit [72] (Leaky ReLU) with a slope of 0.1, and a 2×2 max pooling operation with a stride size of 2 for downsampling. The two-convolutional-layer residual blocks contain two consecutive convolutional layers with a kernel size of 3×3 and a convolutional residual path [73] connecting the in and out tensors of the two convolutional layers. The numbers of the input channels and the output channels at each level of the downsampling path were set to 4, 64, 128, and 256 and 64, 128, 256, and 512, respectively.

Symmetrically, the upsampling path contains four upsampling convolutional blocks with the same design as the downsampling convolutional blocks, except that the 2× downsampling operation was replaced by a 2× bilinear upsampling operation. The input of each upsampling block is the concatenation of the output tensor from the previous block with the corresponding feature maps at the matched level of the downsampling path passing through the attention gated connection. An attention gate consists of three convolutional layers and a sigmoid operation, which outputs an activation weight map highlighting the salient spatial features [71]. The numbers of the input channels and the output channels at each level of the upsampling path were 1024, 1024, 512, and 256 and 1024, 512, 256, and 128, respectively. Following the upsampling path, a two-convolutional layer residual block together with another single convolutional layer reduces the number of channels to 3, matching that of our ground truth images (i.e., 3-channel RGB images). Additionally, a two-convolutional-layer center block was utilized to connect and match the dimensions of the downsampling path and the upsampling path.

The structure of the discriminator network is illustrated in Figure 2(b). An initial block containing one convolutional layer followed by a Leaky ReLU operation first transformed the 3-channel generator output or ground truth image to a 64-channel tensor. Then, five successive two-convolutional-layer residual blocks were added to perform 2× downsampling and expand the channel numbers of each input tensor. The 2× downsampling was enabled by setting the stride size of the second convolutional layer in each block as 2. After passing through the five blocks, the output tensor was averaged and flattened to a one-dimensional vector, which was then fed into two fully connected layers to obtain the probability of the input image being the standard IHC-stained image.

The full image dataset contains 25 WSIs from 19 unique patients, making a set of 20,910 image patches, each with a size of 1024×1024 pixels. For the training of each virtual staining model used in our cross-validation studies, the dataset was divided as follows: (1) Test set: images from the WSIs of 1-2 unique patients (~10%, not overlapped with training or validation patients); after splitting out the test set, the remaining WSIs were further divided to the (2) validation set: images from 2 of the WSIs (~10%), and (3) training set: images from the remaining WSIs (~80%). The network models were optimized using image patches of 256×256 pixels, which were randomly cropped from the images of 1024×1024 pixels in the training dataset. An Adam optimizer with weight decay [74] was used to update the learnable parameters at a learning rate of 1×10^{-4} for the generator network and 1×10^{-5} for the discriminator network, with a batch size of 28. The generator/discriminator update frequency was set to $2:1$. Finally, the best model was selected based on the best MSE loss, assisted with the visual assessment of the validation images. The networks converged after ~120 hours of training.

4.5. Implementation Details. The image preprocessing was implemented in MATLAB using version R2018b (Math-Works). The virtual staining network was implemented using Python version 3.9.0 and Pytorch version 1.9.0. The training was performed on a desktop computer with an Intel Xeon W-2265 central processing unit (CPU), 64 GB random-access memory (RAM), and an Nvidia GeForce GTX 3090 graphics processing unit (GPU).

4.6. Pathologists' Blind Evaluation of HER2 Images. For the evaluation of WSIs, 24 high-resolution WSIs were randomly shuffled, rotated, flipped, and uploaded to an online image viewing platform that was shared with three board-certified pathologists to blindly evaluate and score the HER2 status of each WSI using the Dako HercepTest scoring system [53]. For the evaluation of sub-ROI images, the 240 image patches were randomly shuffled, rotated, flipped, and uploaded to an online image sharing platform GIGAmacro (https://www.gigamacro.com/). These 240 image patches used for staining quality evaluation can be accessed at https://viewer.gigamacro.com/collections/u08mwIpUDAwf R1vQ?sb=date&sd=asc.

The pathologists' blinded assessments are provided in Supplementary Data 1.

4.7. Statistical Analysis. A Chi-square test (two-sided) was performed to compare the agreement of the HER2 scores evaluated based on the virtual staining and the standard IHC staining. Paired *t*-tests (one-sided) were used to compare the image quality of virtual staining vs. standard IHC staining. We first calculated the differences between the scores of the virtual and IHC image patches cropped from the same positions, i.e., subtracted the score of each IHC stained image from the score of the corresponding virtually stained image. Then, one-sided *t*-tests were performed to compare the differences with 0, by each feature metric and each pathologist (see the Supplementary Information). For

all tests, a P value of ≤ 0.05 was considered statistically significant. All the analyses were performed using SAS v9.4 (The SAS Institute, Cary, NC).

4.8. Numerical Evaluation of HER2 Images. For the feature-based quantitative assessment of HER2 images (reported in Figure 6), a color deconvolution [57] was performed to separate the nucleus stain channel (i.e., hematoxylin stain) and the membrane stain channel (i.e., diaminobenzidine stain, DAB), as shown in Supplementary Figure 9. The nucleus segmentation map was obtained using the Otsu's thresholding method [75] followed by morphological operations (e.g., image erosion and image dilation) on the hematoxylin channel. Based on the binary nucleus segmentation map, the number of nuclei and the average nucleus area were extracted by counting the number of connected regions and measuring the average region area. For the evaluation of the membrane stain, the separated DAB image channel was first transformed into the HSV color space. Then, the segmentation map of the membrane stain was obtained by applying a threshold (s) to the saturation channel. By gradually increasing the threshold value (s) from 0.1 to 0.5 with a step size of 0.02, the ratio of the total segmented membrane stain area to the entire image FOV (i.e., 1024×1024 pixels) was calculated, creating the characteristic curve [54] (Supplementary Figure 9). The area under the characteristic curve can be accordingly extracted, providing a robust metric for evaluating HER2 expression levels. By setting the threshold value (s) to 0.25, the ratio of the largest connected component in the membrane segmentation map to the entire image FOV was also extracted as the membrane region connectedness [55].

For the characterization of the color distribution reported in Supplementary Figures 6-7, the nucleus stain channel and the membrane stain channel were split using the same color deconvolution method [57] as in Supplementary Figure 9. For each stain channel, the histogram of all the normalized pixel values was created and followed by a nonparametric kernel-smoothing to fit the distribution profile [76]. y-axes (i.e., the frequency) of the color histograms shown in Supplementary Figures 6-7 were normalized by the total pixel counts.

Data Availability

Data supporting the results demonstrated by this study are available within the main text and the Supplementary Information. The full set of images used for the HER2 status and stain quality assessment studies can be found in the Supplementary Data 1 file and at: https://viewer.gigamacro .com/collections/u08mwIpUDAwfR1vQ?sb=date&sd=asc. The full pathologist reports can be found in the Supplementary Data 1 file. The full statistical analysis report can be found in Supplementary Data 2 file. Raw WSIs corresponding to patient specimens were obtained under UCLA IRB 18-001029 from the UCLA Health private database for the current study and therefore cannot be made publicly available.

Additional Points

Code availability. All the deep-learning models used in this work employ standard libraries and scripts that are publicly available in Pytorch. The codes used in this manuscript can be accessed through GitHub: https://github.com/baibijie/HER2-virtual-staining.

Conflicts of Interest

A.O., Y.R., B.B., H.W., K.d.H, and Y.Z. have a pending patent application related to the work reported in the manuscript.

Authors' Contributions

Y.R. and A.O. conceived the research, B.B., H.W., and N.B.D. conducted the experiments, B.B., H.W., Y.L., Y.W., J.L., and J.Z., developed the codes. B.B., H.W., Y.L., Y.W., X.Z., Y.Z., X.Y., and W.D. processed the data. B.B., H.W., and J.Z. trained the network models. K.d.H and F.C. developed the image-sharing workflow and platform. M.A.D., E.K., and H.S.L. performed diagnosis and stain efficacy assessment on the virtual and histologically stained slides. B.B., H.W., Y.L., M.A.D., E.K., and A.O. prepared the manuscript, and all authors contributed to the manuscript. A.O. supervised the research. Bijie Bai, Hongda Wang and Yuzhu Li contributed equally to this work.

Acknowledgments

Supplementary Materials

The Supplementary Materials include Supplementary Figures 1-9, Supplementary Tables 1-2 and Supplementary Note 1 (IHC HER2 staining protocol). *(Supplementary Materials)*

References

[1] A. H. Coons, H. J. Creech, and R. N. Jones, "Immunological properties of an antibody containing a fluorescent group," *Proceedings of the Society for Experimental Biology and Medicine*, vol. 47, no. 2, pp. 200–202, 1941.

[2] G. Whiteside and R. Munglani, "TUNEL, Hoechst and immunohistochemistry triple-labelling: an improved method for detection of apoptosis in tissue sections—an update," *Brain Research Protocols*, vol. 3, no. 1, pp. 52-53, 1998.

[3] T. Scholzen and J. Gerdes, "The Ki-67 protein: from the known and the unknown," *Journal of Cellular Physiology*, vol. 182, no. 3, pp. 311–322, 2000.

[4] S. Surget, M. P. Khoury, and J.-C. Bourdon, "Uncovering the role of p53 splice variants in human malignancy: a clinical perspective," *Oncotargets and Therapy*, vol. 7, pp. 57–68, 2013.

[5] Z. Mitri, T. Constantine, and R. O'Regan, "The HER2 receptor in breast cancer: pathophysiology, clinical use, and new advances in therapy," *Chemotherapy Research and Practice*, vol. 2012, Article ID 743193, 7 pages, 2012.

[6] J. A. Ramos-Vara and M. A. Miller, "When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry—the red, Brown, and blue technique," *Veterinary Pathology*, vol. 51, no. 1, pp. 42–87, 2014.

[7] J. A. Ramos-Vara, "Technical aspects of immunohistochemistry," *Veterinary Pathology*, vol. 42, no. 4, pp. 405–426, 2005.

[8] M. G. Rojo, G. Bueno, and J. Slodkowska, "Review of imaging solutions for integrated quantitative immunohistochemistry in the pathology daily practice," *Folia Histochemica et Cytobiologica*, vol. 47, no. 3, pp. 349–354, 2009.

[9] Y. Rivenson, H. Wang, Z. Wei et al., "Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue," vol. 3, no. 6, pp. 466–477, 2019.

[10] Y. Zhang, K. de Haan, Y. Rivenson, J. Li, A. Delis, and A. Ozcan, "Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue," *Light: Science & Applications*, vol. 9, no. 1, p. 78, 2020.

[11] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light: Science & Applications*, vol. 8, no. 1, p. 23, 2019.

[12] S. Ryu, N. Martino, S. J. J. Kwok, L. Bernstein, and S. H. Yun, "Label-free histological imaging of tissues using Brillouin light scattering contrast," *Biomedical Optics Express*, vol. 12, no. 3, pp. 1437–1448, 2021.

[13] Z. Chen, W. Yu, I. H. M. Wong, and T. T. W. Wong, "Deep-learning-assisted microscopy with ultraviolet surface excitation for rapid slide-free histological imaging," *Biomedical Optics Express*, vol. 12, no. 9, pp. 5920–5938, 2021.

[14] P. Pradhan, T. Meyer, M. Vieth et al., "Computational tissue staining of non-linear multimodal imaging using supervised and unsupervised deep learning," *Biomedical Optics Express*, vol. 12, no. 4, pp. 2280–2298, 2021.

[15] J. Li, J. Garfinkel, X. Zhang et al., "Biopsy-free in vivo virtual histology of skin using deep learning," *Light: Science & Applications*, vol. 10, no. 1, p. 233, 2021.

[16] Y. Liu, X. Li, A. Zheng et al., "Predict Ki-67 positive cells in H&E-stained images using deep learning independently from IHC-stained images," *Frontiers in Molecular Biosciences*, vol. 7, 2020.

[17] B. He, S. Bukhari, E. Fox et al., "AI-enabled in silico immunohistochemical characterization for Alzheimer's disease," *Cell Reports Methods*, vol. 2, no. 4, article 100191, 2022.

[18] M. Chen, B. Zhang, W. Topatana et al., "Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning," *NPJ Precision Oncology*, vol. 4, pp. 1–7, 2020.

[19] N. Naik, A. Madani, A. Esteva et al., "Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains," *Nature Communications*, vol. 11, no. 1, p. 5727, 2020.

[20] H. D. Couture, L. A. Williams, J. Geradts et al., "Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype*j*," *NPJ Breast Cancer*, vol. 4, pp. 1–8, 2018.

[21] D. Bychkov, N. Linder, A. Tiulpin et al., "Deep learning identifies morphological features in breast cancer predictive of cancer ERBB2 status and trastuzumab treatment efficacy," *Scientific Reports*, vol. 11, no. 1, p. 4037, 2021.

[22] A. Binder, M. Bockmayr, M. Hägele et al., "Morphological and molecular breast cancer profiling through explainable machine learning," *Nature Machine Intelligence*, vol. 3, no. 4, pp. 355–366, 2021.

[23] G. Shamai, Y. Binenbaum, R. Slossberg, I. Duek, Z. Gil, and R. Kimmel, "Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer," *JAMA Network Open*, vol. 2, no. 7, article e197700, 2019.

[24] H. Xu, J. R. Clemenceau, S. Park, J. Choi, S. H. Lee, and T. H. Hwang, "Spatial heterogeneity and organization of tumor mutation burden and immune infiltrates within tumors based on whole slide images correlated with patient survival in bladder cancer," *Journal of Pathology Informatics*, vol. 13, article 554527, 2020.

[25] J. M. Dolezal, A. Trzcinska, C. Y. Liao et al., "Deep learning prediction of BRAF-RAS gene expression signature identifies noninvasive follicular thyroid neoplasms with papillary-like nuclear features," *Modern Pathology*, vol. 34, no. 5, pp. 862–874, 2021.

[26] D. Anand, K. Yashashwi, N. Kumar, S. Rane, P. H. Gann, and A. Sethi, "Weakly supervised learning on unannotated H&E-stained slides predicts BRAF mutation in thyroid cancer with high accuracy," *The Journal of Pathology*, vol. 255, no. 3, pp. 232–242, 2021.

[27] R. H. Kim, S. Nomikou, N. Coudray et al., "A deep learning approach for rapid mutational screening in melanoma," *BioRxiv*, no. article 610311, 2020.

[28] J. L. Connolly, S. J. Schnitt, H. H. Wang, J. A. Longtine, A. Dvorak, and H. F. Dvorak, "Role of the Surgical Pathologist in the Diagnosis and Management of the Cancer Patient," in *Holland-Frei Cancer Medicine*, BC Decker, 6th edition edition, 2003.

[29] I. Rubin and Y. Yarden, "The basic biology of HER2," *Annals of Oncology*, vol. 12, pp. S3–S8, 2001.

[30] N. Iqbal and N. Iqbal, "Human epidermal growth factor receptor 2 (HER2) in cancers: overexpression and therapeutic implications," *Molecular Biology International*, vol. 2014, Article ID 852748, 9 pages, 2014.

[31] J. S. Ross and J. A. Fletcher, "The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy," *The Oncologist*, vol. 3, no. 4, pp. 237–252, 1998.

[32] M. Bilous, M. Dowsett, W. Hanna et al., "Current perspectives on HER2 testing: a review of National Testing Guidelines," *Modern Pathology*, vol. 16, no. 2, pp. 173–182, 2003.

[33] H. J. Burstein, "The distinctive nature of HER2-positive breast cancers," *New England Journal of Medicine*, vol. 353, no. 16, pp. 1652–1654, 2005.

[34] Å. Borg, A. K. Tandon, H. Sigurdsson et al., "HER-2/neu amplification predicts poor survival in node-positive breast cancer," *Cancer Research*, vol. 50, no. 14, pp. 4332–4337, 1990.

[35] R. L. B. Costa and B. J. Czerniecki, "Clinical development of immunotherapies for HER2$^+$ breast cancer: a review of HER2-directed monoclonal antibodies and beyond," *Breast Cancer*, vol. 6, no. 1, pp. 1–11, 2020.

[36] J. Wang and B. Xu, "Targeted therapeutic options and future perspectives for HER2-positive breast cancer," *Signal Transduction and Targeted Therapy*, vol. 4, no. 1, pp. 1–22, 2019.

[37] A. C. Pinto, F. Ades, E. de Azambuja, and M. Piccart-Gebhart, "Trastuzumab for patients with HER2 positive breast cancer: delivery, duration and combination therapies," *The Breast*, vol. 22, pp. S152–S155, 2013.

[38] G. Hudelist, W. J. Köstler, J. Attems et al., "Her-2/neu-triggered intracellular tyrosine kinase activation: in vivo relevance of ligand-independent activation mechanisms and impact upon the efficacy of trastuzumab-based treatment," *British Journal of Cancer*, vol. 89, no. 6, pp. 983–991, 2003.

[39] D. B. Agus, M. S. Gordon, C. Taylor et al., "Phase I clinical study of pertuzumab, a novel HER dimerization inhibitor, in patients with advanced cancer," *Journal of Clinical Oncology*, vol. 23, no. 11, pp. 2534–2543, 2005.

[40] M. Kaushik Tiwari, D. A. Colon-Rios, H. C. R. Tumu et al., "Direct targeting of amplified gene loci for proapoptotic anticancer therapy," *Nature Biotechnology*, vol. 40, no. 3, pp. 325–334, 2022.

[41] J. Ni, S. H. Ramkissoon, S. Xie et al., "Combination inhibition of PI3K and mTORC1 yields durable remissions in mice bearing orthotopic patient-derived xenografts of HER2-positive breast cancer brain metastases," *Nature Medicine*, vol. 22, no. 7, pp. 723–726, 2016.

[42] J. C. Kang, W. Sun, P. Khare et al., "Engineering a HER2-specific antibody–drug conjugate to increase lysosomal delivery and therapeutic efficacy," *Nature Biotechnology*, vol. 37, no. 5, pp. 523–526, 2019.

[43] J. C. Singh, K. Jhaveri, and F. J. Esteva, "HER2-positive advanced breast cancer: optimizing patient outcomes and opportunities for drug development," *British Journal of Cancer*, vol. 111, no. 10, pp. 1888–1898, 2014.

[44] H. Creedon, L. A. Balderstone, M. Muir et al., "Use of a genetically engineered mouse model as a preclinical tool for HER2 breast cancer," *Disease Models & Mechanisms*, vol. 9, no. 2, pp. 131–140, 2016.

[45] E. A. Fry, P. Taneja, and K. Inoue, "Clinical applications of mouse models for breast cancer engaging HER2/neu," *Integrative Cancer Science and Therapeutics*, vol. 3, no. 5, pp. 593–603, 2016.

[46] C. De Giovanni, G. Nicoletti, E. Quaglino et al., "Vaccines against human HER2 prevent mammary carcinoma in mice transgenic for human HER2," *Breast Cancer Research*, vol. 16, no. 1, p. R10, 2014.

[47] M. P. Piechocki, Y.-S. Ho, S. Pilon, and W.-Z. Wei, "Human ErbB-2 (Her-2) transgenic mice: a model system for testing Her-2 based vaccines," *Journal of Immunology*, vol. 171, no. 11, pp. 5787–5794, 2003.

[48] N. V. Jordan, A. Bardia, B. S. Wittner et al., "HER2 expression identifies dynamic functional states within circulating breast cancer cells," *Nature*, vol. 537, no. 7618, pp. 102–106, 2016.

[49] C. Giesen, H. A. O. Wang, D. Schapiro et al., "Highly multiplexed imaging of tumor tissues with subcellular resolution

by mass cytometry," *Nature Methods*, vol. 11, no. 4, pp. 417–422, 2014.

[50] D. R. Glenn, K. Lee, H. Park et al., "Single cell magnetic imaging using a quantum diamond microscope," *Nature Methods*, vol. 12, no. 8, pp. 736–738, 2015.

[51] M. Hafner, M. Niepel, M. Chung, and P. K. Sorger, "Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs," *Nature Methods*, vol. 13, no. 6, pp. 521–527, 2016.

[52] S. Vickovic, G. Eraslan, F. Salmén et al., "High-definition spatial transcriptomics for in situ tissue profiling," *Nature Methods*, vol. 16, no. 10, pp. 987–990, 2019.

[53] "Herceptest™ Interpretation Manual - Breast Cancer," https://www.agilent.com/cs/library/usermanuals/public/28630_herceptest_interpretation_manual-breast_ihc_row.pdf.

[54] R. Mukundan, *A Robust Algorithm for Automated HER2 Scoring in Breast Cancer Histology Slides Using Characteristic Curves. in Medical Image Understanding and Analysis*, M. Valdés Hernández and V. González-Castro, Eds., vol. 723, Springer International Publishing, 2017.

[55] R. Mukundan, "Analysis of image feature characteristics for automated scoring of HER2 in histology slides," *Journal of Imaging*, vol. 5, no. 3, p. 35, 2019.

[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[57] G. Landini, G. Martinelli, and F. Piccinini, "Colour deconvolution: stain unmixing in histological imaging," *Bioinformatics*, vol. 37, no. 10, pp. 1485–1487, 2021.

[58] K. de Haan, Y. Zhang, J. E. Zuckerman et al., "Deep learning-based transformation of H&E stained tissues into special stains," *Nature Communications*, vol. 12, no. 1, p. 4884, 2021.

[59] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Venice, Italy, 2017.

[60] J. R. Lakowicz, H. Szmacinski, K. Nowaczyk, K. W. Berndt, and M. Johnson, "Fluorescence lifetime imaging," *Analytical Biochemistry*, vol. 202, no. 2, pp. 316–330, 1992.

[61] P. I. H. Bastiaens and A. Squire, "Fluorescence lifetime imaging microscopy: spatial resolution of biochemical processes in the cell," *Trends in Cell Biology*, vol. 9, no. 2, pp. 48–52, 1999.

[62] M. E. Andersen and R. Z. Muggli, "Microscopical techniques in the use of the molecular optics laser examiner Raman microprobe," *Analytical Chemistry*, vol. 53, no. 12, pp. 1772–1777, 1981.

[63] B. Bai, H. Wang, Y. Li et al., "Label-free virtual HER2 immunohistochemical staining of breast tissue using deep learning," 2021, https://arxiv.org/abs/2112.05240.

[64] A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, "Computer control of microscopes using μManager," *Current Protocols in Molecular Biology*, vol. 92, no. 1, pp. 14–20, 2010.

[65] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[66] S. Saalfeld, R. Fetter, A. Cardona, and P. Tomancak, "Elastic volume reconstruction from series of ultra-thin microscopy sections," *Nature Methods*, vol. 9, no. 7, pp. 717–720, 2012.

[67] H. Wang, Y. Rivenson, Y. Jin et al., "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, 2019.

[68] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, Hawaii, US, 2017.

[69] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[70] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, pp. 79–82, 2005.

[71] O. Oktay, J. Schlemper, L. L. Folgoc et al., "Attention u-net: Learning where to look for the pancreas," https://arxiv.org/abs/1804.03999.

[72] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *International Conference on Machine Learning, ICML*, vol. 30, no. 1, p. 3, 2013.

[73] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Nevada, US, 2016.

[74] D. P. Kingma and J. A. Ba, "A Method for Stochastic Optimization," 2017, https://arxiv.org/abs/1412.6980.

[75] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[76] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.