BCI: Breast Cancer Immunohistochemical Image Generation through Pyramid Pix2pix

Shengjie Liu[1] Chuang Zhu*[1] Feng Xu*[2] Xinyu Jia[1] Zhongyue Shi[2] Mulan Jin[2]
[1]Beijing University of Posts and Telecommunications, Beijing, China
[2]Capital Medical University, Beijing, China
{shengjie.Liu, czhu, jiaxinyubupt}@bupt.edu.cn
drxufeng@mail.ccmu.edu.cn {shizhongyue815, kinmokuran}@163.com

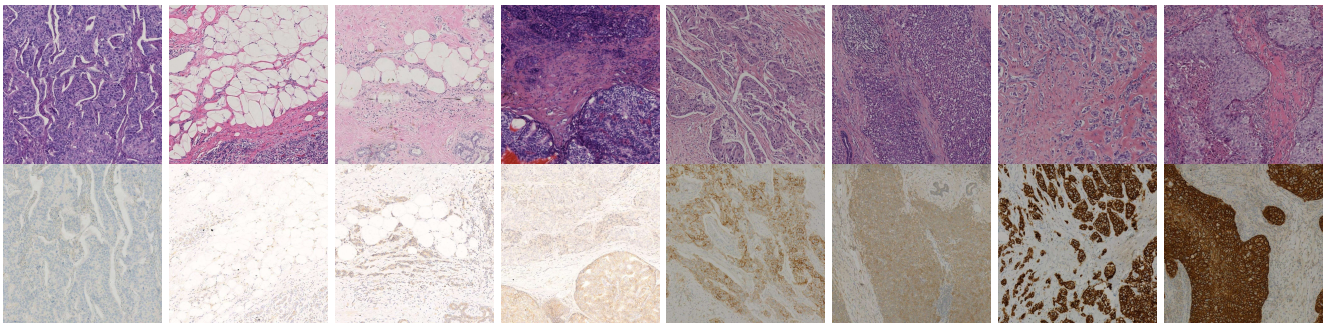


Figure 1. Samples of BCI. **Top**: HE-stained patches. **Bottom**: IHC-stained patches. Each column represents a HE-IHC image pair. It contains four expression levels of HER2 (0, 1+, 2+, 3+).

Abstract

The evaluation of human epidermal growth factor receptor 2 (HER2) expression is essential to formulate a precise treatment for breast cancer. The routine evaluation of HER2 is conducted with immunohistochemical techniques (IHC), which is very expensive. Therefore, for the first time, we propose a breast cancer immunohistochemical (BCI) benchmark attempting to synthesize IHC data directly with the paired hematoxylin and eosin (HE) stained images. The dataset contains 4870 registered image pairs, covering a variety of HER2 expression levels.

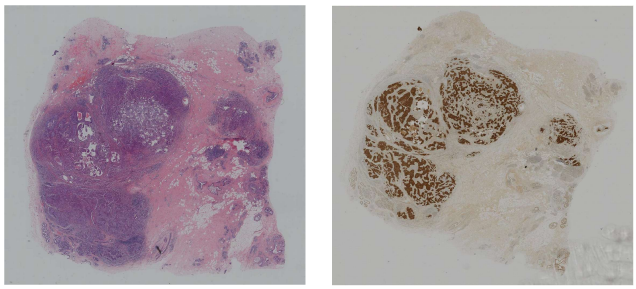
Based on BCI, as a minor contribution, we further build a pyramid pix2pix image generation method, which achieves better HE to IHC translation results than the other current popular algorithms. Extensive experiments demonstrate that BCI poses new challenges to the existing image translation research. Besides, BCI also opens the door for future pathology studies in HER2 expression evaluation based on the synthesized IHC images. BCI dataset can be downloaded from https://bupt-ai-cz.github.io/BCI.

*Corresponding authors: Chuang Zhu (czhu@bupt.edu.cn), Feng Xu (drxufeng@mail.ccmu.edu.cn)

1. Introduction

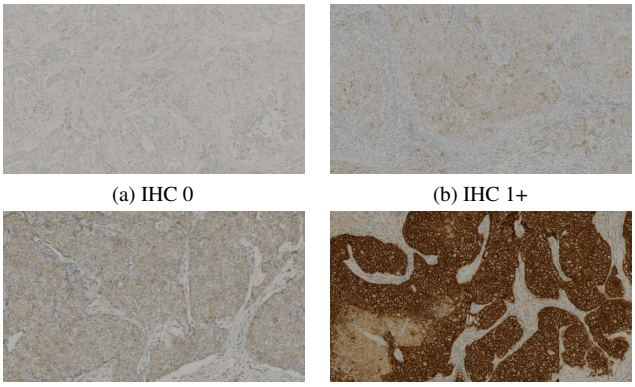
According to work [35], breast cancer is a leading cause of death for women. Accurate diagnosis and therapy are key factors to reduce the mortality rate of breast cancer patient [44]. The histopathological checking is a gold standard to identify breast cancer. To achieve this, the tumor materials are first made into hematoxylin and eosin (HE) stained slices (a slice is shown in Fig. 2(a)). Then, the diagnosis is performed by pathologists through observing the HE slices under the microscope or analyzing the digitized whole slice images (WSI). For diagnosed breast cancer, it is essential to formulate a precise treatment plan by checking the expression of specific proteins, such as human epidermal growth factor receptor 2 (HER2) [16]. The breast cancer with overexpression of HER2 is prone to have aggressive clinical behaviour, and thus accurate therapy should be formulated accordingly.

The routine evaluation of HER2 expression is conducted with immunohistochemical techniques (IHC) [11]. Specifically, one additional IHC-stained slice (a slice is shown in Fig. 2(b)) is first prepared. Then the pathologists will check the IHC-stained slice to obtain the HER2 expression status: IHC 0, no staining is observed or membrane staining that is incomplete and is faint/barely perceptible and in $\leq 10\%$ of tumor cells (Fig. 3(a)); IHC 1+, incomplete



(a) An example of HE slice. (b) An example of IHC slice.

Figure 2. Visualization of HE-stained and IHC-stained slices.



(a) IHC 0 (b) IHC 1+

(c) IHC 2+ (d) IHC 3+

Figure 3. Visualization of four kinds of HER2 expressions.

membrane staining that is faint/barely perceptible and in >10% of tumor cells (Fig. 3(b)); IHC 2+, weak to moderate complete membrane staining observed in >10% of tumor cells (Fig. 3(c)); IHC 3+, circumferential membrane staining that is complete, intense, and in >10% of tumor cells (Fig. 3(d)) [39]. The detection of HER2 expression is critical to the formulation of follow-up treatment plans for breast cancer. However, it is very expensive to conduct HER2 evaluation through the additional preparation of IHC-stained slice. Then the question is can we synthesize the IHC-stained image based on HE-stained WSI? In case of success, we can conduct HER2 expression evaluation directly based on the synthesized IHC-stained slices.

This paper presents the above challenge for the first time, and tries to solve it through image-to-image translation technique. Image translation aims to learn the mapping between an input source-domain image and an output target-domain image [7]. In recent years, some methods and datasets have been proposed to promote the research of image-to-image translation.

Pix2pix [9] proposes a universal translation method for paired images. Since then, there have been other supervised image translation algorithms based on pix2pix that can be applied to specific scenes: pix2pixHD [36] has achieved very good results in high resolution paired image translation; work [28] proposes an enhanced pix2pix optimized for image dehazing. Besides, there are also many excellent methods [5, 8, 21, 22] for unsupervised image translation inspired by these pioneering works [13, 42, 45].

Dataset is the key factor for image translation, especially for supervised methods. Many fields have proposed datasets pertinently. However, there are only a few works for the medical image translation applications. RegGAN [15] implements a general image translation method for both paired and unpaired images on BraTS [25] dataset. In the field of breast cancer, a few of datasets such as BCNB [40] have been proposed for automatic diagnosing, however, there are no datasets for HE to IHC staining for HER2 detection. This task requires structural level aligned datasets, which poses great challenges due to the difficulty of the acquisition of well paired HE-IHC images. To the best of our knowledge, there are no public image translation datasets exists for HER2 detection in breast cancer tissue.

To spur research in this area, we introduce BCI, a structural aligned dataset for the translation of HE-stained slices to immunohistochemical results (Fig. 1). We also propose a method optimized for this task. We benchmark several state-of-the-art (SOTA) algorithms for image translation tasks. In summary, this paper makes the following contributions:

- We collect and build BCI: a paired HE to HER2 expression image translation dataset. To our knowledge, BCI is the first large-scale publicly available dataset for immunohistochemical image generation.
- We propose a pyramid pix2pix method to generate immunohistochemical image based on HE. Compared with other pix2pix-like methods [9, 36], our method can constrain the generated image at multiple scales and achieve better results on our BCI dataset.
- We conduct extensive experiments on BCI and LLVIP dataset [10] to explore the gains that different scales bring to the model, which demonstrate the flexibility and versatility of multi-scale constraints.

2. Related Work

In this section, we will review some image translation algorithms, as well as some datasets that are often used in image translation tasks.

2.1. Image Translation

Image translation algorithm establishes a mapping between two domain images. It is often used in image semantic, image synthesis, and image super-resolution, etc. Image

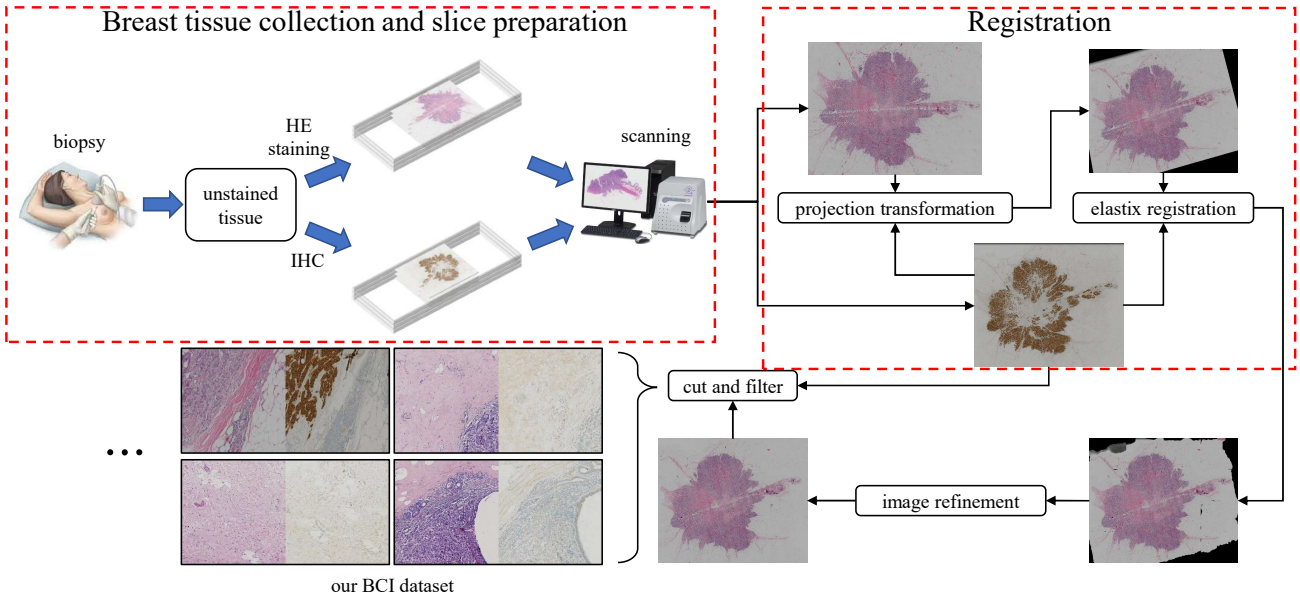


Figure 4. The establishment of our BCI dataset is generally divided into three steps: 1) breast tissue collection and slice preparation; 2) registration of images in the two domains; 3) post-processing including image refinement and patches cutting.

translation algorithms can be divided into unsupervised image translation and supervised image translation.

Unsupervised image translation does not require aligned datasets, which makes the application range of this method very wide. Many works [5,8,13,21,22,42,45] are dedicated to the translation of unpaired images. This type of methods can randomly extract images from two domains during training, which can achieve better image style transfer when it is difficult to obtain paired data. In addition to solving the problem of image translation from one domain to another, there are some methods [3,4] that creatively solve the image translation between multiple domains.

However, unsupervised methods also have a certain limitation: for paired datasets, it may not be possible to establish an accurate mapping between the two domains. To overcome this problem, work [41] adds additional per-patch labels (e.g. background, necrosis, fibrosis, etc.) to CycleGAN during training. For the classification of patches, huge amounts of human work are still required. Therefore, supervised image translation methods still have great application value. Pix2pix [9] is a pioneering supervised image translation algorithm. It is a general image translation algorithm that can be applied to various image translation tasks. In addition to the adversarial loss between the generator and the discriminator, it also calculates the pixel-level difference between the generated image and the ground truth to continuously improve the generator's effect. Pix2pixHD [36] optimizes the generator structure on the basis of pix2pix to make the generation of high-resolution images better. EPDN [28] follows the overall structure of pix2pix, which uses a multi-resolution generator, a multi-scale discriminator, and an enhancer for image dehazing tasks. SRGAN [19] and ESRGAN [37] apply the generative adversarial network to image super-resolution tasks, which are essentially a kind of image translation. There are also some supervised methods [27, 34] that are widely used in semantic image synthesis. These models take semantic information as input and translate it into real images.

In the medical field, image translation already has some applications. RegGAN [15] proposes a general image translation model, which adds a U-net structure registration network after the generator in pix2pix. It calculates the loss between the output image of the registration network and the ground truth, which makes RegGAN achieve good results in both paired and unpaired data. In the field of pathological images, there are some works that [2, 33] translate non-standard stained sections into standard stained sections, providing new ideas for the normalization of pathological image staining.

2.2. Datesets

The datasets used for image translation tasks are abundant, and these datasets can also be divided into two categories: paired and unpaired. Many paired datasets [1, 6, 29–31, 43] can be used for translation between semantic distribution maps and real images; Cityscapes [6] and Foggy Cityscapes [32] can be used for image dehazing research; CelebAMask-HQ [20] and FFHQ-Aging [26] are two large-scale face datasets, are used in the research of generating face images from segmentation masks. Part

of the images in work [17] can be used for conversion between day and night. LLVIP [10] contains registered images in two domains of visible light and infrared light, which can be used for translation between visible light images and infrared light images. BraTS [25] is a dataset in the medical field, in which T1 weighted images and T2 weighted images can be used for the translation of brain MRI images. Selfie2anime [12] is an unpaired dataset that provides images in two domains, selfies and cartoon characters, which can be used for the research of transforming real pictures into cartoon styles. There are also some multi-domain datasets such as AFHQ [4] and RaFD [18], these datasets can be used for unsupervised image translation and image synthesis in multiple domains. Note that all paired datasets can be used to train unsupervised image translation models.

3. BCI Dataset

The application of deep learning in the medical field has attracted more and more attention. There are already some brain image translation datasets to promote research on brain science, however, there is still no relevant data for pathological image translation. Therefore, we propose the BCI dataset, in order to better promote the research of pathological image translation. We hope that BCI can play a positive role in the diagnosis of breast cancer. At the same time, as a benchmark, our dataset can help analyze the advantages and disadvantages of current image translation algorithms.

Our overall process of building the dataset is shown in Fig. 4. Next, we will introduce more details about this dataset.

3.1. Collection

The data scanning equipment is Hamamatsu NanoZommer S60, a pathology section scanner with a scanning speed of 60 seconds per slice. The scanning resolution of the equipment is 0.46 μm per pixel. We scanned more than 600 pathological slices of breast cancer tissues and sorted out the WSI stained with HE and the corresponding immunohistochemical WSI of 319 breast cancer patients. In the image registration process, we filter out WSI pairs that are unable to complete the alignment. Finally, we got 4870 pairs of HE-IHC patches from 51 different WSI image pairs.

3.2. Registration

For a piece of pathological tissue, the doctor will cut two tissue samples from it for HE staining and HER2 detection. Therefore, there will be differences in the morphology of the two pathological samples. Besides, the tissue samples will be stretched or squeezed to a certain extent during slice preparation, which will increase the difference between the samples. In order to make the images of the two domains aligned, we need to perform registration processing on the images.

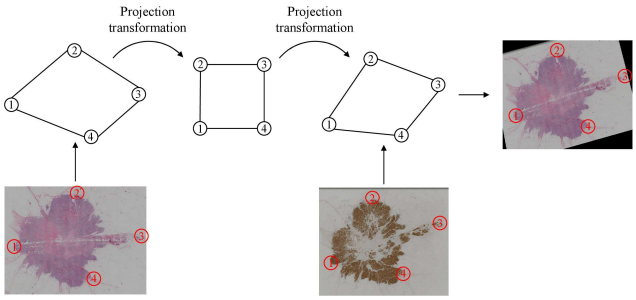


Figure 5. Projection transformation by manually selecting corresponding points. The HE image can be initially aligned with the IHC image after a two-step projection transformation.

Projection transformation. First, we use the method of human-computer interaction projection transformation to roughly align the WSI pairs. This method requires manually selecting no less than 4 pairs of corresponding points on the two WSIs (Fig. 5), then through the method of projection mapping, the irregular quadrilateral determined by the four points on the HE WSI is first mapped to a square, and then the square is mapped to the irregular quadrilateral determined by the four points on the IHC WSI. In this process, the HE image is basically aligned with the contour of the corresponding IHC image by translation and rotation, and the resolution of the two images is kept consistent. At this time, there are still some deviations inside the HE and IHC images, and further registration is required.

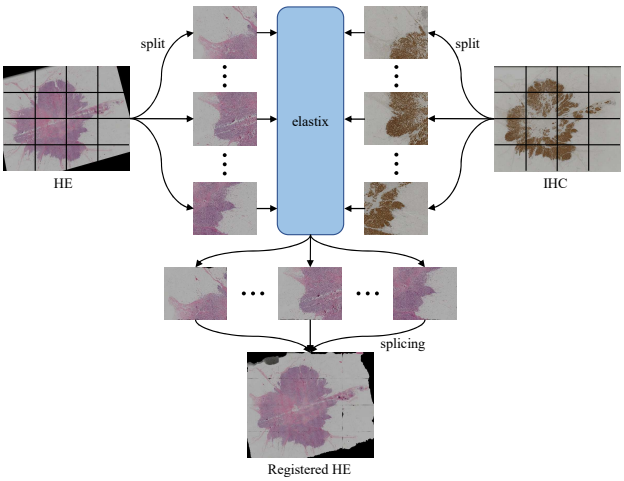


Figure 6. The process of elastix registration. The roughly aligned HE and IHC images are divided into blocks, and each block is registered separately with elastix. Finally, the registered blocks are re-spliced

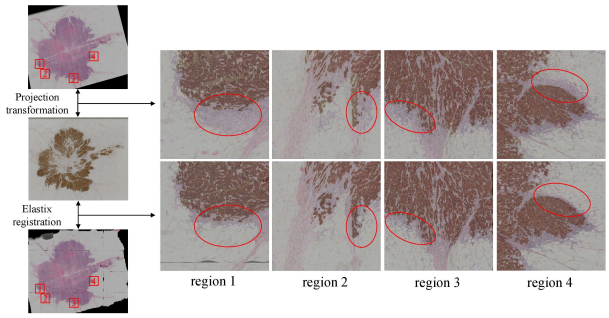


Figure 7. By overlapping the registration result with the corresponding IHC image, the difference between projection transformation and elastix registration can be seen: projection transformation can only roughly overlap the two images, but cannot achieve detailed registration; after elastix registration, the overlap in details is realized.

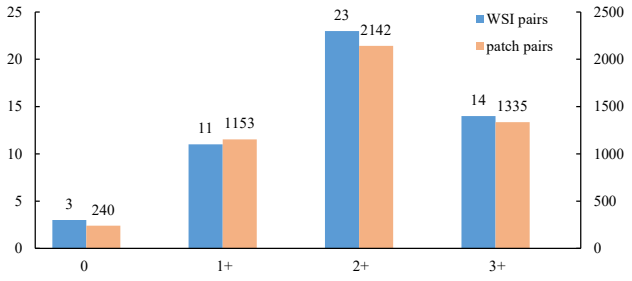


Figure 8. Image statistics of IHC results.

Elastix registration. Second, we use the registration toolbox elastix [14] to perform fine-grained regional non-rigid registration. This process can align the details of the two domain images as much as possible. Fig. 7 shows the detailed alignment of elastix registration based on projection transformation. For each WSI, because its resolution is too high (about 20,000 pixels on a side), the computational power and time consumed by direct registration are huge, in order to improve the efficiency of registration, we divide it into 16 blocks for registration respectively. At the same time, since elastix cannot directly process the RGB image, we split the image and use only a single channel for registration, save the transformation file of the channel, and then apply the transformation file to the other two channels. Finally, we merge the registered images of the three channels and then stitch each registered block into WSI. The registration process of elastix is shown in Fig. 6.

3.3. Post-processing

Our block registration method is efficient, however, during the registration process, the expansion and contraction of the image will leave a gap on the edge of each image block. Therefore, we need to remove the black border between the blocks and fill it with the surrounding content. Finally, the registered WSI image is cut into 1024×1024 size patches. Finally, we will filter out blank and not well-aligned areas.

Our BCI dataset contains 4870 pairs of pediatric pathological image patches with a resolution of 1024×1024. These patches are from the WSIs of 51 patients. The immunohistochemical results of these 51 patients included four categories: 0, 1+, 2+, and 3+. Fig. 8 shows the distribution of the 51 WSIs and the number of patches from different IHC results.

4. Proposed Method

4.1. Architecture

Our BCI dataset presents a new challenge for image translation. In our dataset, the images of the two domains are paired and registered at the structural level. However, due to the existence of image differences between the two domains, some positions cannot achieve pixel-level alignment, which makes the existing pix2pix series of algorithms difficult to work; at the same time, we need to perform targeted output for each HE stained image, which is not the strength of the unsupervised algorithms. Therefore, we propose a pyramid pix2pix model suitable for structural aligned data. Our overall framework is shown in Fig. 9.

The L_1 loss in pix2pix algorithm directly calculates the difference between the generated image and the ground truth, which is too restrictive on the generated image. For our BCI dataset, we need to weaken the constraints of L_1 loss, while aligning the generated image and ground truth at other scales. Inspired by scale-space theory [24], we will perform the same scale transformation on the generated image and ground truth. The scale transformation consists of two steps: 1) Using a low-pass filter to smooth the image. 2) Downsampling the smooth image. Since the Gaussian kernel is the only linear kernel that realizes the image scale transformation, our low-pass filter uniformly uses the Gaussian kernel with a standard deviation of 1. With the progress of Gaussian filtering, the image becomes more and more blurred, and we reduce the resolution by downsampling to remove redundant pixels. For each resolution level (octave), multiple Gaussian convolutions are performed to achieve scale transformation. Our pix2pix pyramid has several octaves, the first layer of each octave is obtained by downsampling the last image of the previous octave; each octave has 5 layers and performs 4 Gaussian blurrings. For each output of octave, we define it as a scale (Fig. 10). In our Gaussian Pyramid, we extract the first layer of images in each octave to calculate the loss. The loss for each scale is denoted as $S_i (i = 1, 2, 3 \cdots)$:

$$S_i = \mathbb{E}_{x,y,z}[\|F_i(y) - F_i(G(x,z))\|_1], \quad (1)$$

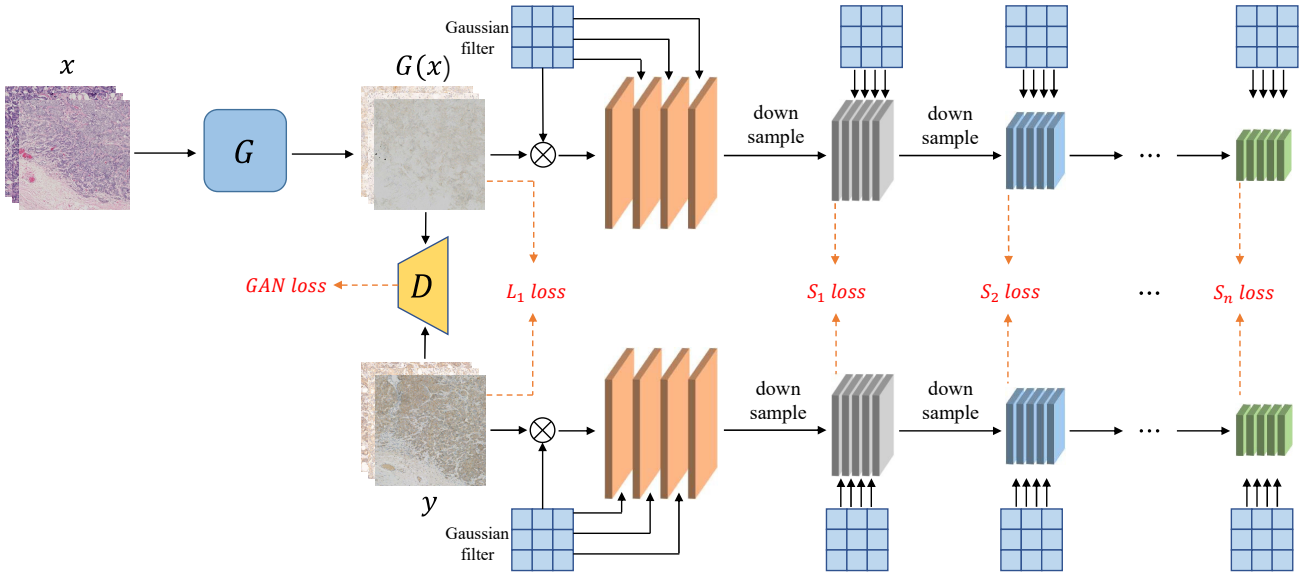
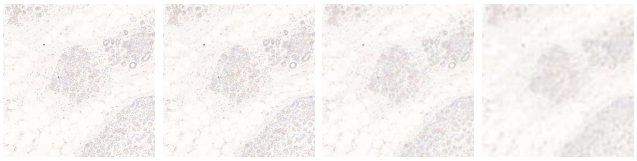


Figure 9. The framework of the proposed pyramid pix2pix. The image of each scale is obtained from the image of the previous scale after four Gaussian convolutions and one downsampling.



(a) scale 1 (b) scale 2 (c) scale 3 (d) scale 4

Figure 10. Visualization of a sample image at different scales. Gaussian convolution makes the image gradually blurred.

where F_i and G represent the Gaussian filtering operation and the generator, respectively. x, y and z represent the input image, the ground truth, and random noise, respectively. Even if we cannot make the generated image highly consistent with the ground truth in the first octave, we can still make the generated image close to the ground truth on a higher-dimensional scale. Our multi-scale loss is recorded as:

$$L_{multi-scale} = \sum_i \lambda_i S_i, \tag{2}$$

where λ_i represents the weight of scale i. Our adversarial loss is still consistent with pix2pix:

$$\begin{aligned} L_{cGAN}(G,D) =& \mathbb{E}_{x,y}[logD(x,y)]+ \\ & \mathbb{E}_{x,z}[log(1-D(x,G(x,z)))], \end{aligned} \tag{3}$$

where generator G tries to minimize this function while discriminator D tries to maximize it. We still keep the L_1 loss in pix2pix in order to maintain the constraints on the original resolution:

$$L_1 = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1]. \tag{4}$$

At this point, our overall objective function is:

$$G^* = arg \min_G \max_D L_{cGAN}(G,D) + \lambda_1 L_1 + L_{multi-scale}. \tag{5}$$

5. Experiments

In this section, we will use several image translation algorithms to conduct experiments on our BCI dataset. Our experiment was performed on NVIDIA Tesla T4 16GB GPU.

5.1. Implementation

In the pix2pix algorithm, we tried two generator structures, unet256 and resnet-9blocks. Through experiments, we found that resnet-9blocks obviously has a better generation effect. Therefore, in our proposed method, we use the generator structure of resnet-9blocks as the baseline, while the discriminator structure uses the default patchGAN; the Gaussian kernel used is 3×3 with a standard deviation of 1; the input images are not preprocessed; the batch size is set to 2; the optimizer used is Adam; the total number of training epochs is set to 100: the learning rate of first 50 epochs is set to 0.0002 and the learning rate of the remaining 50 epochs gradually drops to 0. In pix2pixHD, both the generator and the discriminator adopt the default settings, and the image preprocessing, the number of training epochs and the optimization strategy are consistent with those of pix2pix. In the cycleGAN algorithm, due to memory limitations, we randomly crop the image to 512×512 resolution before training, and other settings are also consistent with pix2pix.

5.2. Metrics

We use Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) as the evaluation indicators for the quality of the generated image. PSNR is based on the error between the corresponding pixels of two images and is the most widely used objective evaluation index. However, the evaluation result of PSNR may be different from the evaluation result of the Human Visual System (HVS). Therefore, we also use SSIM [23, 38], which comprehensively measures the differences in image brightness, contrast, and structure. This evaluation result is closer to the human visual system.

5.3. Benchmark Results

Method	PSNR(dB)	SSIM
cycleGAN	16.203	0.373
pix2pix(unet generator)	18.654	0.419
pix2pix(resnet generator)	19.328	0.440
pix2pixHD	19.634	0.471
ours	**21.160**	**0.477**

Table 1. Comparison of PSNR and SSIM values of different methods on BCI dataset.

Method	PSNR(dB)	SSIM
cycleGAN	11.22	0.214
pix2pix(unet generator)	10.769	0.176
pix2pix(resnet generator)	12.082	0.207
pix2pixHD	11.156	0.228
ours	**12.191**	**0.278**

Table 2. Comparison of PSNR and SSIM values of different methods on LLVIP dataset.

For the unsupervised method, we choose the most representative cycleGAN. It can be seen from the experimental results (Fig. 11 and Table 1) that as an unsupervised image translation algorithm, cycleGAN cannot establish an accurate mapping from HE to IHC results. For these registered image pairs, it can only achieve "style" migration, but it is completely impossible to identify the cancer areas. As a representative algorithm of supervised image translation, pix2pix with a resnet generator can basically stain the cancerous area. Its PSNR and SSIM indicators are significantly higher than cycleGAN, but the quality of the generated image is poor, pix2pix with unet generator is even worse. Besides, the staining effect of pix2pix generated images is quite different from the correct results of IHC, especially in the areas where HER2 is highly expressed. Pix2pixHD uses a two-stage generator structure and performs adversarial discriminating on multiple scales. Its high-resolution image generation quality is slightly better than pix2pix on the whole, therefore, it has higher PSNR and SSIM than pix2pix. However, in some areas with low HER2 expression, pix2pixHD may incorrectly generate dark browns.

The result of our method is better than pix2pix and pix2pixHD in terms of authenticity. In the identification of HER2 expression, our method is better in the case of low expression of HER2 (0/1+), the difference between the generated image and ground truth is slight (Fig. 11(a)(b)); when the expression level of HER2 is 2+, the image we generate will be lighter than ground truth, but the effect is still better than other methods (Fig. 11(c)); when HER2 is highly expressed (3+), our method is the same as other methods, unable to identify areas of high expression of HER2 (Fig. 11(d)), which is also a major issue that needs to be resolved in the future. It is still very challenging to establish an accurate mapping from HE to HER2 expression on our dataset. We still need to explore more effective methods to improve the accuracy of the translation.

On the LLVIP[1] dataset, our method also achieves the best PSNR and SSIM (Table 2), which proves that our method is not only suitable for the translation of pathological images but also has a certain versatility.

5.4. Multi-scale Analysis

Configuration	PSNR(dB)	SSIM
pix2pix	19.328	0.440
pix2pix+S1 (ours)	**21.160**	**0.477**
pix2pix+S1+S2 (ours)	21.033	0.469
pix2pix+S1+S2+S3 (ours)	21.138	0.472

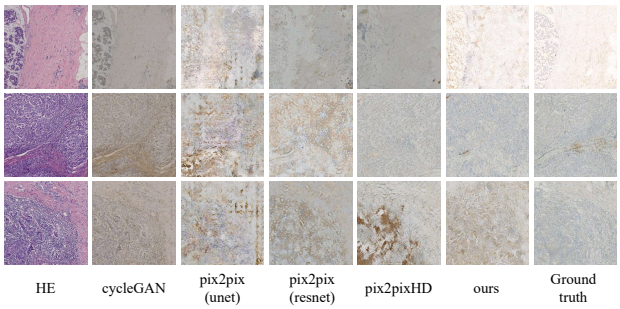
Table 3. Multi-scale analysis on BCI dataset.

Configuration	PSNR(dB)	SSIM
pix2pix	12.082	0.207
pix2pix+S1 (ours)	12.189	**0.279**
pix2pix+S1+S2 (ours)	12.173	0.277
pix2pix+S1+S2+S3 (ours)	**12.191**	0.278

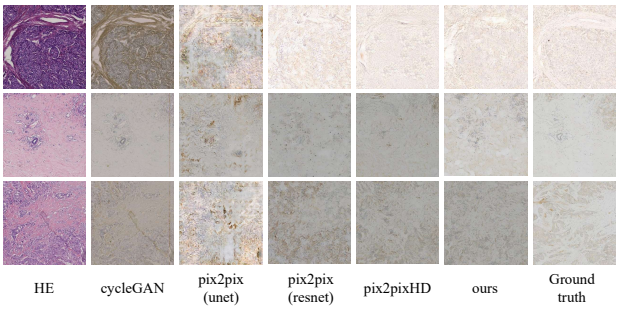
Table 4. Multi-scale analysis on LLVIP dataset.

Our pyramid pix2pix has the flexibility to change the number of pyramid layers to accommodate different datasets. On our BCI dataset, we tried the gains of different pyramid levels. Table 3 shows that the model with a two-layer pyramid structure (pix2pix+S1) achieves the

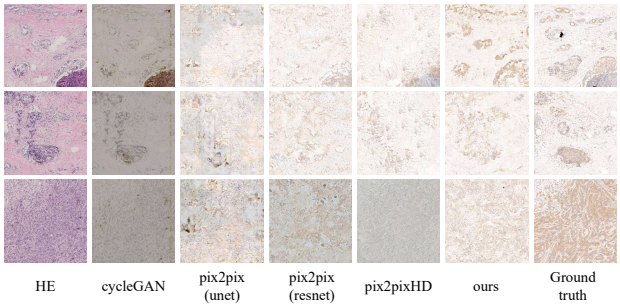
[1]Visit the link https://bupt-ai-cz.github.io/LLVIP for details of the LLVIP dataset



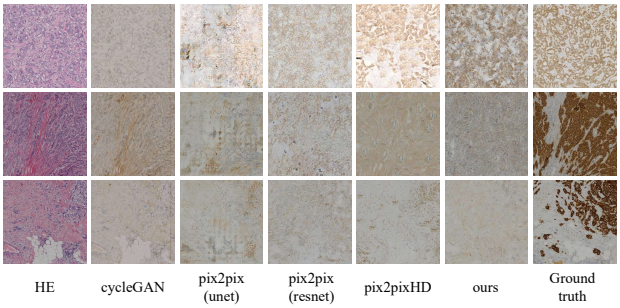
(a) Visualization of different methods on IHC 0 images. The image we generated is very close to the ground truth.



(b) Visualization of different methods on IHC 1+ images. The image we generated is very close to the ground truth.



(c) Visualization of different methods on IHC 2+ images. In general, our generated image has a lighter color than ground truth, however, it is still better than other methods.



(d) Visualization of different methods on IHC 3+ images. In this case, all methods are difficult to accurately identify the cancer area, which is a huge challenge.

Figure 11. Visualization of different methods on different HER2 expressions.

highest PSNR and SSIM, which demonstrates that it is more reasonable to constrain the generated images and ground truth on the second scale. By optimizing the loss function of scale two (S1), the model effect can be greatly improved. On LLVIP dataset, a four-layer pyramid pix2pix can achieve the approximate effect of a two-layer model (Table 4), which shows that the constraint of the high level also improves the generation effect compared to pix2pix.

5.5. Subjective Validation

In addition to objective metrics, we also invited two pathologists to diagnose HER2 expression of the generated images. To avoid the influence of subjective factors, we randomly selected 40 real-generated IHC pairs, and shuffled the order of these 80 images. A generated IHC image is considered accurate if the generated image and its corresponding real image are diagnosed at the same level. The accuracy of these 40 generated images is shown in Table 5. The results show that there is still a long way to go for current methods before clinical application, which also proves the importance of the BCI dataset in further research.

6. Conclusion

In this paper, we propose BCI, a new dataset in the field of pathology images for the translation of HE stained breast tissue section to its IHC results. This task puts forward new requirements for image translation algorithms, which is to accurately identify the expression area and expression level of HER2 while ensuring the authenticity of the generated image. In addition, we also propose pyramid pix2pix, an image-to-image translation model suitable for registered image pairs.

	pathologist1	pathologist2
Accuracy(%)	37.5	40.0

Table 5. Accuracy of generated IHC images.

It is still very challenging to establish an accurate mapping from HE to IHC results. There is still a need for more effective methods to improve the accuracy of the translation. In addition, in the future, we will explore the difference in HER2 evaluation between synthetic IHC images and real IHC images. Then we will further study the possibility of formulating accurate clinical treatment plans for breast cancer using synthetic IHC images.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 62176167, and in part by the BUPT innovation and entrepreneurship support program under Grant 2022-YC-T046.

References

[1] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Cocostuff: Thing and stuff classes in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1209–1218, 2018. 3

[2] Hyungjoo Cho, Sungbin Lim, Gunho Choi, and Hyunseok Min. Neural stain-style transfer learning using gan for histopathological images. *arXiv preprint arXiv:1710.08543*, 2017. 3

[3] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018. 3

[4] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8188–8197, 2020. 3, 4

[5] Min Jin Chong and David Forsyth. Gans n'roses: Stable, controllable, diverse image to image translation (works for videos too!). *arXiv preprint arXiv:2106.06561*, 2021. 2, 3

[6] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 3

[7] Shihua Huang, Cheng He, and Ran Cheng. Multimodal image-to-image translation via a single generative adversarial network. *arXiv preprint arXiv:2008.01681*, 2020. 2

[8] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018. 2, 3

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 2, 3

[10] Xinyu Jia, Chuang Zhu, Minzhen Li, Wenqi Tang, and Wenli Zhou. Llvip: A visible-infrared paired dataset for low-light vision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3496–3504, 2021. 2, 4

[11] Fariba Damband Khameneh, Salar Razavi, and Mustafa Kamasak. Automated segmentation of cell membranes to evaluate her2 status in whole slide images using a modified deep learning network. *Computers in biology and medicine*, 110:164–174, 2019. 1

[12] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv preprint arXiv:1907.10830*, 2019. 4

[13] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *International Conference on Machine Learning*, pages 1857–1865. PMLR, 2017. 2, 3

[14] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2009. 5

[15] Lingke Kong, Chenyu Lian, Detian Huang, Zhenjiang Li, Yanle Hu, and Qichao Zhou. Breaking the dilemma of medical image-to-image translation. *arXiv preprint arXiv:2110.06465*, 2021. 2, 3

[16] David La Barbera, António Polónia, Kevin Roitero, Eduardo Conde-Sousa, and Vincenzo Della Mea. Detection of her2 from haematoxylin-eosin slides through a cascade of deep learning classifiers via multi-instance learning. *Journal of Imaging*, 6(9):82, 2020. 1

[17] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on graphics (TOG)*, 33(4):1–11, 2014. 4

[18] Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel HJ Wigboldus, Skyler T Hawk, and AD Van Knippenberg. Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377–1388, 2010. 4

[19] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 3

[20] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5549–5558, 2020. 3

[21] Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Singh, and Ming-Hsuan Yang. Drit++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision*, 128(10):2402–2417, 2020. 2, 3

[22] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in neural information processing systems*, pages 700–708, 2017. 2, 3

[23] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999. 7

[24] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 5

[25] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 2, 4

[26] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *European Conference on Computer Vision*, pages 739–755. Springer, 2020. 3

[27] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019. 3

[28] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, 2019. 2, 3

[29] Radim Šára Radim Tyleček. Spatial pattern templates for recognition of objects with regular structure. In *Proc. GCPR*, Saarbrucken, Germany, 2013. 3

[30] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *European conference on computer vision*, pages 102–118. Springer, 2016. 3

[31] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016. 3

[32] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018. 3

[33] M Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni. Staingan: Stain style transfer for digital histological images. In *2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019)*, pages 953–956. IEEE, 2019. 3

[34] Vadim Sushko, Edgar Schönfeld, Dan Zhang, Juergen Gall, Bernt Schiele, and Anna Khoreva. You only need adversarial supervision for semantic image synthesis. *arXiv preprint arXiv:2012.04781*, 2020. 3

[35] Lindsey A Torre, Farhad Islami, Rebecca L Siegel, Elizabeth M Ward, and Ahmedin Jemal. Global cancer in women: burden and trends. *Cancer Epidemiology and Prevention Biomarkers*, 26(4):444–457, 2017. 1

[36] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018. 2, 3

[37] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 3

[38] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 7

[39] Antonio C Wolff, M Elizabeth Hale Hammond, Kimberly H Allison, Brittany E Harvey, Pamela B Mangu, John MS Bartlett, Michael Bilous, Ian O Ellis, Patrick Fitzgibbons, Wedad Hanna, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline focused update. *Archives of pathology & laboratory medicine*, 142(11):1364–1382, 2018. 2

[40] Feng Xu, Chuang Zhu, Wenqi Tang, Ying Wang, Yu Zhang, Jie Li, Hongchuan Jiang, Zhongyue Shi, Jun Liu, and Mulan Jin. Predicting axillary lymph node metastasis in early breast cancer using deep learning on primary tumor biopsy slides. *Frontiers in Oncology*, page 4133, 2021. 2

[41] Zhaoyang Xu, Carlos Fernández Moro, Béla Bozóky, and Qianni Zhang. Gan-based virtual re-staining: a promising solution for whole slide image analysis. *arXiv preprint arXiv:1901.04059*, 2019. 3

[42] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857, 2017. 2, 3

[43] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017. 3

[44] Chuang Zhu, Fangzhou Song, Ying Wang, Huihui Dong, Yao Guo, and Jun Liu. Breast cancer histopathology image classification through assembling multiple compact cnns. *BMC medical informatics and decision making*, 19(1):1–17, 2019. 1

[45] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2, 3