

Region-guided CycleGANs for Stain Transfer in Whole Slide Images

Joseph Boyd¹, Irène Villa², Marie-Christine Mathieu², Eric Deutsch², Nikos Paragios³, Maria Vakalopoulou^{*1}, and Stergios Christodoulidis^{*1}

¹ MICS Laboratory, CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France `firstname.lastname@centralesupelec.fr`

² Gustave Roussy Cancer Campus, 94800 Villejuif, France
`firstname.lastname@gustaveroussy.fr`

³ Therapanacea, 75014 Paris, France
`n.paragios@therapanacea.eu`

Abstract. In whole slide imaging, commonly used staining techniques based on hematoxylin and eosin (H&E) and immunohistochemistry (IHC) stains accentuate different aspects of the tissue landscape. In the case of detecting metastases, IHC provides a distinct readout that is readily interpretable by pathologists. IHC, however, is a more expensive approach and not available at all medical centers. Virtually generating IHC images from H&E using deep neural networks thus becomes an attractive alternative. Deep generative models such as CycleGANs learn a semantically-consistent mapping between two image domains, while emulating the textural properties of each domain. They are therefore a suitable choice for stain transfer applications. However, they remain fully unsupervised, and possess no mechanism for enforcing biological consistency in stain transfer. In this paper, we propose an extension to CycleGANs in the form of a region of interest discriminator. This allows the CycleGAN to learn from unpaired datasets where, in addition, there is a partial annotation of objects for which one wishes to enforce consistency. We present a use case on whole slide images, where an IHC stain provides an experimentally generated signal for metastatic cells. We demonstrate the superiority of our approach over prior art in stain transfer on histopathology tiles over two datasets. Our code and model are available at <https://github.com/jcboyd/miccai2022-roigan>.

Keywords: Stain transfer · CycleGANs · Region-based discriminator

1 Introduction

The use of histopathological whole slide images (WSI) is considered the gold standard for the diagnosis and prognosis of cancer patients. These slides contain biopsies of pathological tissue from patients and are typically stained in order to highlight different tissue structures. One of the most common tissue staining

* These authors contributed equally to this work.

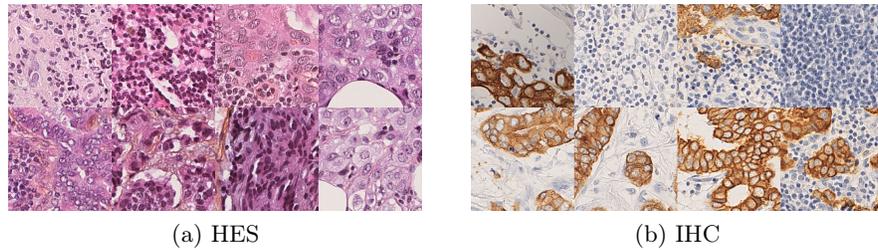


Fig. 1: Stain transfer translates between unpaired HES (a) and IHC (b) tiles (Gustave Roussy dataset). Cancer cells visible in IHC by the golden DAB stain.

protocols is hematoxylin and eosin staining (H&E), sometimes augmented with saffron (HES) (Figure 1a). Under the H&E protocol, the hematoxylin component stains cell nuclei with a purple-blue color, while the eosin component stains the extracellular matrix and cytoplasm in pink [20]. Another common staining is based on immunohistochemistry (IHC) (Figure 1b). This staining involves the process of selectively identifying proteins in cells and highlighting them using a chromogen. One of the most common IHC stains is based on hematoxylin together with the diaminobenzidine (DAB) chromogen forming an H-DAB stain. Under this protocol, different staining configurations can be used, targeting different cell proteins (e.g., Ki67, HER2, hormone receptors). The resulting stained tissue can provide a distinct readout, for example with AE1/AE3, in which DAB localises on cancer cell membranes. In the case of breast cancer, the presence of metastatic cells in the axillary lymph nodes can greatly affect the prognosis of a patient. Lymph node status, assessed through the WSI inspection of dissected sentinel or axillary lymph nodes, is therefore a crucial diagnostic readout. In cases where the diagnosis is difficult to make (e.g. micro-metastasis or isolated tumor cells) IHC-stained slides that highlight cancer cells are necessary.

With the recent advances in deep learning, there is a growing interest in automatically processing digitised slides. We present a methodology to computationally transfer from H&E to IHC WSIs. Our main contribution is a region-based discriminator network within a CycleGAN framework, which utilises automatically extracted regions of interest to improve stain localisation. Unlike fully supervised schemes, our method learns from unpaired H&E and IHC slides and produces a stain transfer that can serve as a soft segmentation for metastasis detection. Experiments on two datasets illustrate the success of our method, while demonstrating that baseline models are unable to reliably localise the DAB stain and, consequently, cancer cells. To the best of our knowledge, this is the first work to propose such a region-guided CycleGAN, as well as the first to provide robust models for stain transfer on the WSI level. Synthesised IHC slides show the potential of our method as a clinical visualisation tool, as well as in metastasis segmentation pipelines for diagnosis.

2 Related Work

Generative adversarial networks (GANs) are a popular family of generative models in histopathology image analysis [21]. Various types of GANs have been applied to histopathology images for stain transfer [22], stain normalisation [18,2,1], cell segmentation [7], data augmentation [4] and representation learning [3]. A widely used variant of GANs are image-to-image translation networks, such as `pix2pix` [23], which learns to translate between image *pairs*, and CycleGANs [9], which relax the pairing assumption to enable *unpaired* and fully unsupervised image-to-image translation. CycleGANs have been used for stain transfer before, notably in [22], which augments the CycleGAN training criterion with additional priors, so as to guide an otherwise unsupervised model.

Region of interest (RoI) information is one possible enrichment to a generative learning task, in conjunction with attention mechanisms. In [19], a CoGAN-inspired setup [14] is used for joint generation of global and RoI images. In [17], pedestrians are edited into predefined scenes using `pix2pix`, with spatial pyramid pooling in the discriminator for direct scrutiny. Pre-trained R-CNN object detection systems have been used to propose regions during GAN training in [8], or as a feature extractor in object-driven GANs [12]. In contrast, we modify the GAN discriminator itself, and leverage automatically derived RoI data.

3 Method

CycleGANs [23] lend themselves to the task of stain transfer between unpaired histopathology tiles. These are unsupervised models incorporating two GAN generators, $G_{XY} : X \rightarrow Y$ and $G_{YX} : Y \rightarrow X$ for *unpaired* image domains X and Y . In an application of stain transfer between two stains, each stain represents a different domain. Although both generators are trained, in our case only one direction of transfer is desired (HES \rightarrow IHC), and the other generator may be discarded after training. CycleGAN training is performed according to,

$$\min_{G_{XY}, G_{YX}} \max_{D_X, D_Y} \mathcal{L}_{CG} = \mathcal{L}_{GAN}(D_Y, G, X, Y) + \mathcal{L}_{GAN}(G_{XY}, D_Y, X, Y) \quad (1)$$

$$+ \lambda_{CYC} \mathcal{L}_{CYC}(G_{XY}, G_{YX}, X, Y) + \lambda_{ID} \mathcal{L}_{ID}(G_{XY}, G_{YX}, X, Y),$$

which combines least squares adversarial losses [16] \mathcal{L}_{GAN} , with PatchGAN discriminators D_X and D_Y , along with a cycle-consistency loss \mathcal{L}_{CYC} to maintain pixel-wise consistency back and forth between domains, identity function losses \mathcal{L}_{ID} for stability, and λ_{CYC} and λ_{ID} are hand-tuned weights.

3.1 Region of interest discrimination

Image photo-realism is enforced in a CycleGAN by a ‘‘PatchGAN’’ discriminator,

$$D_{patch} : X \rightarrow \{0, 1\}^{h \times w}, \quad (2)$$

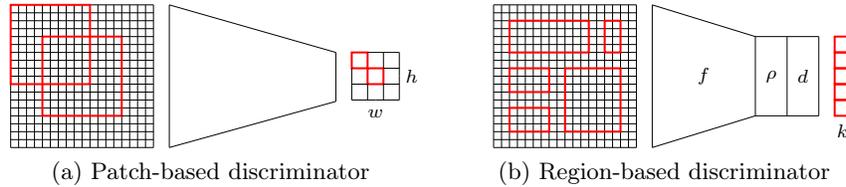


Fig. 2: Conventional PatchGAN discriminates over a regular grid of overlapping receptive fields (a) while a region-based discriminator may discriminate an image on arbitrarily-sized regions (b).

which consists of a sequence of strided convolutions producing a $h \times w$ grid of outputs, with each grid element discriminating an overlapping patch of the inputs, as in Figure 2a. The output grid is compared element-wise with a grid of ground truth labels. As will be quantified in Section 5, this purely unsupervised, standard CycleGAN setup fails to correctly localise DAB in a stain transfer application. This motivates a new type of discriminator, based on region of interest discrimination via a region of interest alignment (RoIAlign) layer. The RoIAlign layer was originally proposed in the Mask R-CNN object detection system [5], and is a generalisation of the classical MaxPool layer, and a way of quantising regions of activation maps of arbitrary size into a standardised dimension. Our proposed RoI discriminator consists of feature extraction layers f followed by a RoIAlign layer ρ , and a final discrimination layer d , as shown in Figure 2b. We formalise it as,

$$D_{roi} : X \times B \rightarrow \{0, 1\}^k, \quad (3)$$

for image domain X and bounding box domain B , where k is the number of bounding boxes. The adversarial loss thus becomes,

$$\mathcal{L}_{ROI}(G, D_{roi}, X, Y) = \frac{1}{2} \mathbb{E}_{\mathbf{x} \in X} [(D_{roi}(\mathbf{x}, \mathcal{B}(\mathbf{x})) - 1)^2] + \frac{1}{2} \mathbb{E}_{\mathbf{y} \in Y} [D_{roi}(G(\mathbf{y}), \mathcal{B}(\mathbf{y}))^2] \quad (4)$$

that is, a least squares adversarial loss, where the operator \mathcal{B} returns the set of bounding boxes of an image, provided by a previously generated object library. A pair of RoI discriminators (one for each domain) can then be trained alongside—or instead of—the PatchGAN discriminators (see Appendix A).

3.2 Library generation for region-based discrimination

To train our proposed region-based discriminator, we build a bounding box library by localising cancer and normal cells using an automatic image processing pipeline (see Appendix B). All input tiles (see Section 4) are first decomposed into their hematoxylin-eosin-DAB components. The expert annotations available for H&E slides identify cancerous regions that may nevertheless contain many

interspersed healthy cells. Blob detection applied directly on the HES images is thus prone to false positives, precisely when precision is more important to our downstream algorithm than recall. To mitigate this effect, we apply a Laplacian of Gaussian (LoG) filter ($\sigma = \{1, 2\}$). We then construct a graph of healthy cells where connectivity is determined by a manual threshold (25 pixels) on Euclidean distance between detected cells. Graph connected components constitute healthy cell clusters, and the convex hull of each cluster is zeroed-out of the expert annotation. In a final step, a second pass of blob detection ($\sigma = \{10, \dots, 12\}$) is applied to detect the larger, dimmer cancer cells in the unmasked regions. In the case of IHC tiles, we instead perform an Otsu threshold on the DAB channel to isolate the cancerous sub-regions. Cancer cell nuclei feature as elliptical discontinuities in the binarised DAB stain, and can again be detected using a LoG filter ($\sigma = \{7, \dots, 10\}$). Smaller blobs ($\sigma = \{1, \dots, 3\}$) outside the segmented region are taken to be normal cells. The library, consisting of over 100k cell bounding boxes, is then used for the training of the RoI discriminator.

3.3 Implementation and training details

In each experiment, the two CycleGAN generators follow the architecture of a baseline approach [22], itself based on a network proposed for style transfer [10]. The tile inputs are of size 256×256 with three (RGB) channels. The generators consist first of strided convolutions to lower the resolution. A sequence of residual blocks [6] of tunable length is then applied. We benchmark 6 and 9 residual blocks. Following this, fractionally-strided convolutions restore the input resolution. The PatchGAN discriminators consist of a standard sequence of five strided convolutions and an output layer. This reduces the input to a 8×8 patch-wise prediction. This architecture, hereafter referred to as CycleGAN 8×8 , in reference to the discriminator output size, is trained according to Equation 1. The baselines and proposed model are all modifications of this core baseline.

Given that our proposed discriminators have a selective receptive field, we control for the receptive field size of the PatchGAN discriminators by modifying the stride of their convolutions. We modify the stride of the first layer from 2 to 1 in baseline CycleGAN 16×16 , doubling the patch output size, and the first two layers in baseline CycleGAN 32×32 , doubling again. We also implement the conditional CycleGAN model [22]. This introduces two additional networks of its own, which are used to classify tiles in each domain. Here, Equation 1 is supplemented with classification and cycle classification losses for the new networks, as well as novel photo-realism and structural similarity losses. Exceptionally, due to the speed and memory constraints of the photo-realism loss, we train this baseline with batch size 2. In addition, we test a plain conditional CycleGAN without these additional losses.

For our proposed model, the RoI discriminator resembles the PatchGAN, with four stride-2 feature convolutions f , followed by a RoIAlign layer ρ , and a final discrimination layer d . Without loss of generality, we restrict cell bounding box dimension to 48×48 , centered on the library nuclei. Although the cell populations in each tile are often imbalanced, we sample balanced numbers of

positives and negatives with replacement, whenever available, equaling $k = 8$ bounding boxes per tile, and we skip tiles containing no library cells. All model weights are randomly initialised and all models use the Adam optimiser [11] with maximum learning rate 2-e4 and $(\beta_0, \beta_1) = (0.5, 0.999)$. Models were trained for 20 epochs with a batch size of 8 randomly sampled tiles, consuming roughly two hours of processing time per model on an NVIDIA Tesla V100 GPU.

4 Datasets

To benchmark our method, we conducted experiments on two breast cancer datasets. One dataset was provided by the Gustave Roussy (GR) Institute consisting of 205 WSIs, corresponding to sentinel lymph node sections for patients of breast cancer. Each lymph node was imaged with both HES and cytokeratin AE1/AE3 IHC. Although the two stainings are performed on closely situated sections of tissue, the lack of precise alignment implies an unpaired image dataset. Metastatic regions of the HES slides were annotated with bounding contours by two expert pathologists. The dataset comprises cases of micro- and macro-metastases (resp. 0.2mm-2mm and ≥ 2 mm tumours), as well as negative cases. Non-overlapping tiles of size 256×256 pixels were extracted at a magnification of 20x from the segmented tissue regions of WSIs using CLAM [15]. For HES slides, the expert annotation is used to extract balanced samples of positive (metastatic tissue) and negative (normal tissue) tiles whereas, for IHC, the thresholded DAB stain substitutes as an experimentally-generated annotation.

CAMELYON16 [13] consists of 1399 hematoxylin-eosin-stained (H&E) slides. Among these, 111 slides were annotated in a fashion similar to the GR dataset. We processed and sampled tiles from these slides in the same way, yielding a dataset of 12000 tiles. Note that this dataset features a slightly different imaging modality, and furthermore contains no IHC slides. Nevertheless, we found that IHC tiles from the GR dataset were fully compatible with CAMELYON16 during model training (provided the imaging magnification), demonstrating that one can combine datasets of different origin for stain transfer.

5 Experimental results

5.1 Tile-level quantitative results

Our primary means of evaluating model performance is with the expert H&E annotations. This provides a good first approximation to the locations of the cancer cells, and where the DAB should appear once the stain transfer has been performed, even though annotation contours often bisect healthy tissue regions or otherwise contain normal cells. For each of our trained generators, we first produce the IHC stain for each of a set of 500 held-out test H&E tiles. We visualise samples of these in Figure 3 for both the proposed model and the baseline CycleGAN 8×8 , alongside the annotation. One may observe the accuracy of our proposed model in localising the DAB stain. Although we observe

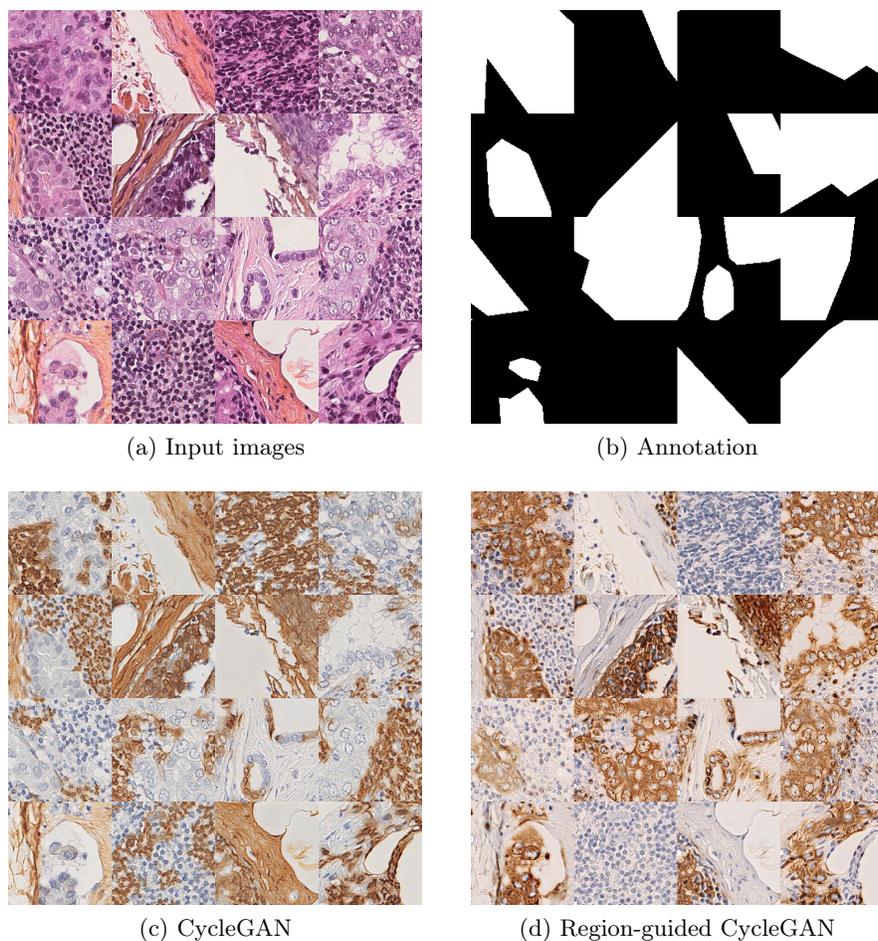


Fig. 3: Sample H&E tiles from the GR dataset (a); with corresponding ground truth annotation masks (b); baseline CycleGAN stain transfer (c); and proposed region-guided stain transfer (d).

a weak correlation between positive tiles and the presence of DAB, the baseline CycleGAN systematically misplaces the DAB stain. This localisation problem was observed across all competing models. We hypothesise our proposed model profits from the object-level supervision, with our proposed RoI discriminator performing discrimination directly centered on cells.

In pursuing a quantitative evaluation, we obtain a binary mask for each synthesised DAB stain in the same manner as was performed in Section 3.2 for real IHC tiles. This mask is then compared with the ground truth annotation by DICE similarity and balanced accuracy (BAC), and we report the mean and

Model	GR Dataset		CAMEYLON16	
	DICE	BAC	DICE	BAC
CycleGAN (8×8)	0.442 ± 0.370	0.490 ± 0.103	0.415 ± 0.372	0.501 ± 0.091
CycleGAN (16×16)	0.365 ± 0.330	0.520 ± 0.125	0.388 ± 0.349	0.502 ± 0.091
CycleGAN (32×32)	0.384 ± 0.324	0.511 ± 0.130	0.403 ± 0.358	0.503 ± 0.100
Conditional CycleGAN	0.341 ± 0.328	0.488 ± 0.120	0.412 ± 0.366	0.501 ± 0.100
Xu et al. (2019) [22]	0.325 ± 0.310	0.474 ± 0.107	0.260 ± 0.257	0.498 ± 0.115
Proposed	0.536 ± 0.360	0.634 ± 0.221	0.533 ± 0.371	0.615 ± 0.197

Table 1: DICE similarity and balanced accuracy (BAC) for all baselines and proposed model for GR dataset and CAMELYON16. Each entry reads mean \pm std. Best results in bold.

standard deviation for each metric across all test tiles in Table 1. We emphasise the ground truth is approximate, as it is intended to capture metastatic regions at a macro level, and often does not exclude healthy sub-regions. For each model, we select the best performing instance in a grid search over the number of generator residual blocks (6, 9), and $\lambda_{cyc} = \{1, 10, 30\}$. Models generally performed better with 6 residual blocks, and with the standard $\lambda_{CYC} = 10$. Our proposed model performs significantly better for $\lambda_{CYC} = 30$, however, seemingly to compensate for the weight of the additional discriminators. We nevertheless note the superiority of our proposed model over all baselines, with 0.536 DICE and 0.634 BAC on the GR dataset, and 0.533 DICE and 0.615 BAC on CAMELYON16. Among the baselines, we observe that changing the discriminator outputs (16×16 and 8×8) can have a positive effect on BAC, but ultimately falls short on DICE. On the other hand, the conditional CycleGANs do not show improvement over the plain baselines. Finally, we observe that our high performance is sustained on CAMELYON16, even though in the absence of a IHC data, GR IHC tiles, of different clinical origin, were used as surrogate.

5.2 Slide-level qualitative results

We further explore the capabilities of our method by generating whole slide outputs, as shown for a macrometastatic case (Appendix C, Figure 3a, b, and c) and negative case (Appendix C, Figure 3d, e, and f) from the GR dataset. One may observe that the HES and IHC ground truth slides are unregistered. Here, the model is applied tile-by-tile and the outputs are recombined to produce slide thumbnails. Inference times were 22s for Appendix C Figure 3b and 14s for Appendix C Figure 3e. We observe the overall consistency of our model outputs with the IHC ground truth. However, false positives, indicated by misplaced DAB signal, remain a problem for our model, and attenuating these will be the subject of future work. Surprisingly, though these test slides belong to the GR dataset, we found the model trained on CAMELYON16 data produced outputs (pictured) at least comparable to those from the GR model, indicating a readily transferable method.

6 Discussion

In this paper we have demonstrated a systematic localisation problem with unsupervised CycleGANs for stain transfer in histopathology tiles, and proposed an improved method using a region-based discriminator, leveraging a library of cell interest regions. We have further shown how datasets of different clinical origin may be successfully combined for learning stain transfer models. The proposed pipeline is a semi-automatic means for extracting additional supervision “for free”, greatly improving the unsupervised baseline. Although for H&E data, we still rely on an expert annotation, our experiments revealed exciting possibilities for further automation. Firstly, due to a parsimonious design (only assumptions about cell size and clustering are made), the library building pipeline was equally applicable to both datasets and would likely generalise to others. Secondly, we found datasets could be combined and that a model trained on CAMELYON16 H&E transfers well to the GR dataset. In the latter case, only annotations from CAMELYON16 have been used, implying reusability of a library once computed on CAMELYON16, a free resource.

Acknowledgments This work was partially supported by the ANR Hagnodice ANR-21-CE45-0007 and ARC SIGNIT201801286. Experiments have been conducted using HPC resources from the “Mésocentre” computing center of CentraleSupélec and École Normale Supérieure Paris-Saclay supported by CNRS and Région Île-de-France.

References

1. de Bel, T., Hermsen, M., Kers, J., van der Laak, J., Litjens, G.: Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology. In: International Conference on Medical Imaging with Deep Learning–Full Paper Track (2018) 3
2. BenTaieb, A., Hamarneh, G.: Adversarial stain transfer for histopathology image analysis. *IEEE transactions on medical imaging* **37**(3), 792–802 (2017) 3
3. Boyd, J., Liashuha, M., Deutsch, E., Paragios, N., Christodoulidis, S., Vakalopoulou, M.: Self-supervised representation learning using visual field expansion on digital pathology. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 639–647 (2021) 3
4. Claudio Quiros, A., Murray-Smith, R., Yuan, K.: Pathologygan: Learning deep representations of cancer tissue. *Journal of Machine Learning for Biomedical Imaging* **2021**(4), 1–48 (2021) 3
5. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017) 4
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 5
7. Hou, L., Agarwal, A., Samaras, D., Kurc, T.M., Gupta, R.R., Saltz, J.H.: Robust histopathology image analysis: To label or to synthesize? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8533–8542 (2019) 3

8. Huang, W., Xu, Y., Oppermann, I.: Realistic image generation using region-phrase attention. arXiv preprint arXiv:1902.05395 (2019) [3](#)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017) [3](#)
10. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016) [5](#)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) [6](#)
12. Li, W., Zhang, P., Zhang, L., Huang, Q., He, X., Lyu, S., Gao, J.: Object-driven text-to-image synthesis via adversarial training. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 12174–12182 (2019) [3](#)
13. Litjens, G., Bandi, P., Ehteshami Bejnordi, B., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., et al.: 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. *GigaScience* **7**(6), giy065 (2018) [6](#)
14. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in neural information processing systems. pp. 469–477 (2016) [3](#)
15. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering* **5**(6), 555–570 (2021) [6](#)
16. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2794–2802 (2017) [3](#)
17. Ouyang, X., Cheng, Y., Jiang, Y., Li, C.L., Zhou, P.: Pedestrian-synthesisgan: Generating pedestrian data in real scene and beyond. arXiv preprint arXiv:1804.02047 (2018) [3](#)
18. Rana, A., Yauney, G., Lowe, A., Shah, P.: Computational histological staining and destaining of prostate core biopsy rgb images with generative adversarial neural networks. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 828–834. IEEE (2018) [3](#)
19. Savioli, N., Vieira, M.S., Lamata, P., Montana, G.: A generative adversarial model for right ventricle segmentation. arXiv preprint arXiv:1810.03969 (2018) [3](#)
20. Suvarna, K.S., Layton, C., Bancroft, J.D.: Bancroft’s theory and practice of histological techniques E-Book. Elsevier health sciences (2018) [2](#)
21. Tschuchnig, M.E., Oostingh, G.J., Gadermayr, M.: Generative adversarial networks in digital pathology: a survey on trends and future potential. *Patterns* **1**(6), 100089 (2020) [3](#)
22. Xu, Z., Moro, C.F., Bozóky, B., Zhang, Q.: Gan-based virtual re-staining: a promising solution for whole slide image analysis. arXiv preprint arXiv:1901.04059 (2019) [3](#), [5](#), [8](#)
23. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017) [3](#)