

Stain Normalization in Digital Histopathology Images

Isabel Amalia Jepsen

DTU Compute, Technical University of Denmark

Introduction

What is a digital histopathology image?

Every time a patient enters the hospital with the suspicion of cancer, a **biopsy** or tissue examination is made. This biopsy is taken from the tissue where the cancer could be placed, e.g. the breast or colon. To examine the tissue, the process given in figure 2 should be completed. A biopsy is taken from the site of interest (a), the tissue is fixed (b), and afterwards sliced into thin slices and mounted on a glass surface (c) and stained to give the tissue color (d). Afterwards the stained tissue can be examined and analyzed on a computer (e) giving information about the diagnosis and the further treatment of the patient (f).

Why do stain variations occur?

As shown in figure 2, the process from biopsy to result is long and not very automated. This means that in each step, the bioanalytic has many ways to solve the same task. E.g. the type of fixation, the thickness of the slice, the type of scanner used to obtain the image and most importantly the staining procedure. **Hematoxylin and eosin** is the most common type of staining used worldwide. However, no standard protocol is followed or quality assurance is made, when using this stain. An example of stain variations is given in figure 1, where the same tissue is stained at four different Danish hospitals.

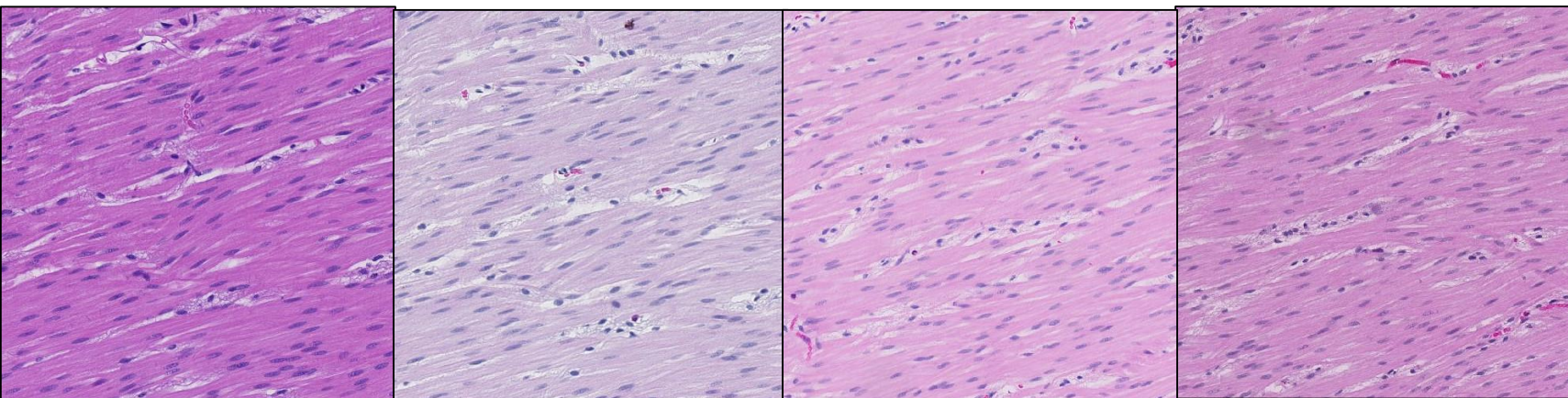


Figure 1: Stain variation resulting after staining the same tissue in four different Danish hospitals

Hypothesis

Why is stain normalization necessary?

When automated image analysis of the tissue is made, the result is highly dependent of the stain intensities given in that specific image. Therefore, stain variations make it difficult for the algorithm to give consistent results across stain variations. Therefore a stain normalization algorithm is suggested, to normalize all images in a batch according to a given target image, which then is expected to increase the performance of automated image analysis tasks.

Method

Data

The data set consists of 100 H&E stained images of **colorectal adenocarcinomas** obtained from a study conducted in the Department of Computer Science, University of Warwick. The images represent different tissue areas with **variations in staining**, artefacts and failed auto focusing, representing outliers normally found in real scenarios. The center position of almost **30.000 nuclei** are manually annotated by an experienced pathologist.

Hematoxylin, Eosin and Background classification

The stain normalization algorithm is based on classifying each pixel into one of the three classes: Hematoxylin (H), Eosin (E) and Background (B). After classifying each pixel into one of those classes, the color and intensity distribution is transformed such that it matches the color and intensity distribution of the target image.

Validation

The goal is to increase the performance of image analysis algorithms, in this project a **nuclei detection algorithm**. The stain normalization algorithm is therefore validated by applying a nuclei detection algorithm on the normalized images, and comparing with the performance of the nuclei detection algorithm on non-normalized images. In figure 3 an image is given without normalization (a), with manually marked centers which represents the ground truth (b) and with the segmented nuclei detected by the nuclei detection algorithm (c).

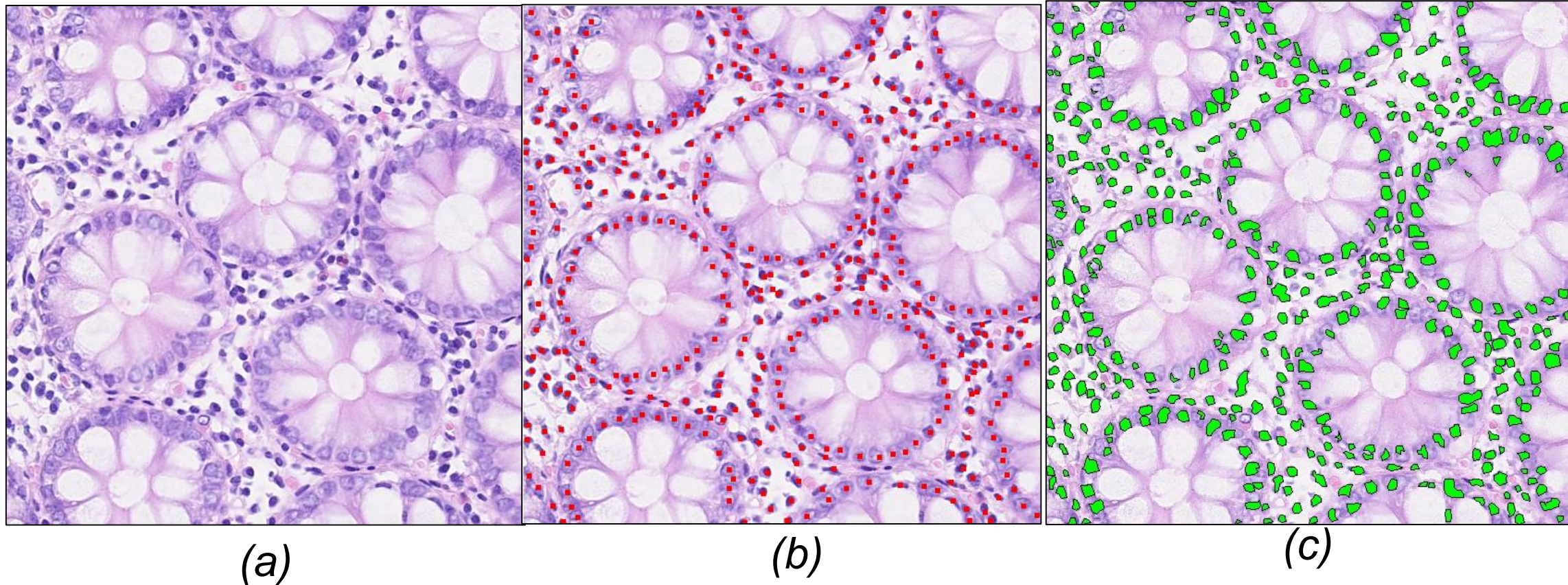


Figure 3: Original image (a), with nuclei manually marked (b), with nuclei detected by algorithm (c).

Results

The results of applying to stain normalization algorithm is given in figure 5. The images before and after normalization are given. The target image is shown in figure 4, which has the color distributions that should be matched.

The F1 score represents the nuclei detected compared with the manually annotation and is wished to be close to 1 to give a maximal performance. This is observed to increase on the images after normalization.

Test	Precision	Recall	F1 score
No normalization	0.54	0.71	0.61
With normalization	0.58	0.74	0.63

Conclusion

The stain normalization algorithm increases the performance of the nuclei detection algorithm. Further work seeks in evaluating the impact of stain normalizing other data sets and evaluating the performance of other image analysis algorithms.

Bibliography

- 1) Image: 2438, Ron Leishman-2438, <http://toonclips.com/designs/2438>
- 2) Image: https://www.google.dk/search?q=tissue+fixation&source=lnms&tbn=isch&sa=X&ved=0ahUKEwIVgbfg7d_TAhVDJCwKH_e0ICfAQ_AUICigB&biw=1680&bih=944#tbn=isch&q=tissue+fixation+paraffin&imgsrc=NbclU3gM5zuj0M:
- 3) Image: https://www.google.dk/search?q=glass+slide+tissue&source=lnms&tbn=isch&sa=X&ved=0ahUKEwjOv7q07d_TAhXFlw_KHeuFCoQQ_AUICigB&biw=1680&bih=944#imgsrc=jle_k1SpJslCpM:
- 4) Image: https://www.google.dk/search?q=glass+slide+tissue&source=lnms&tbn=isch&sa=X&ved=0ahUKEwjOv7q07d_TAhXFlw_KHeuFCoQQ_AUICigB&biw=1680&bih=944#tbn=isch&q=staining+tissue&imgsrc=uKAF8ez4nMXMbM:
- 5) Image: https://www.google.dk/search?q=glass+slide+tissue&source=lnms&tbn=isch&sa=X&ved=0ahUKEwjOv7q07d_TAhXFlw_KHeuFCoQQ_AUICigB&biw=1680&bih=944#tbn=isch&q=computer+histopathology&imgsrc=pyg4KPNsi-3mgM:
- 6) Image: 7834, Ron Leishman-7834, <http://toonclips.com/designs/7834>

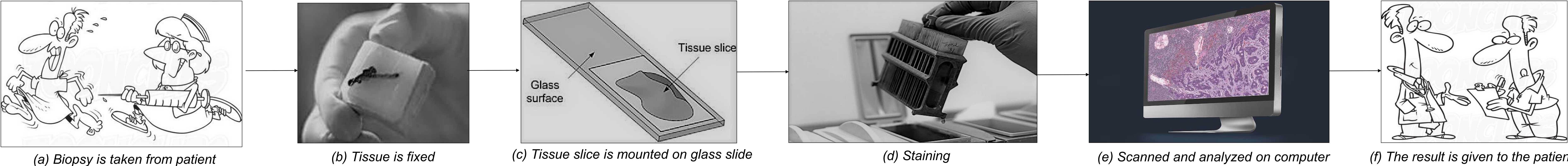
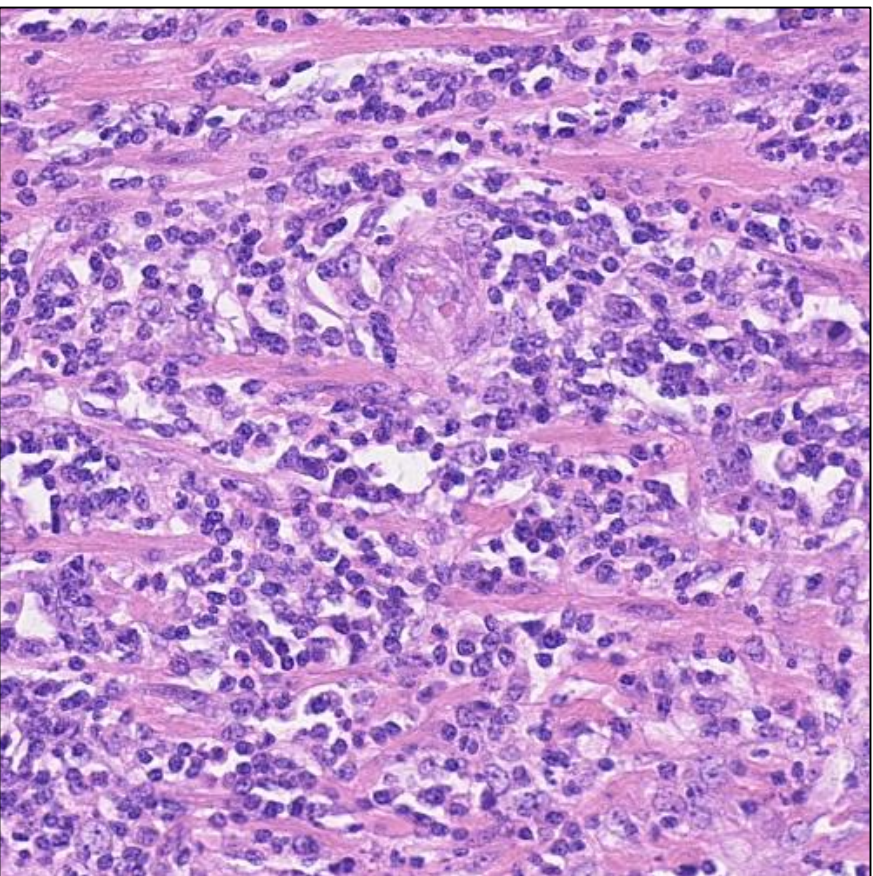
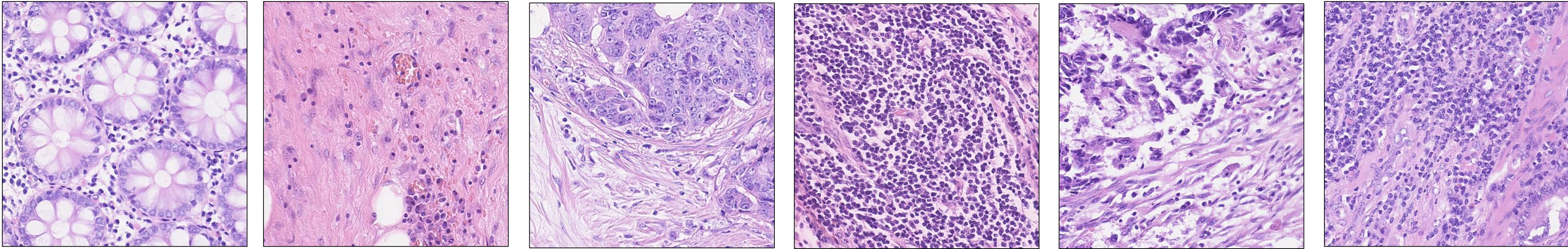


Figure 2: The process from biopsy to result [1]-[6]

Target



Before



After

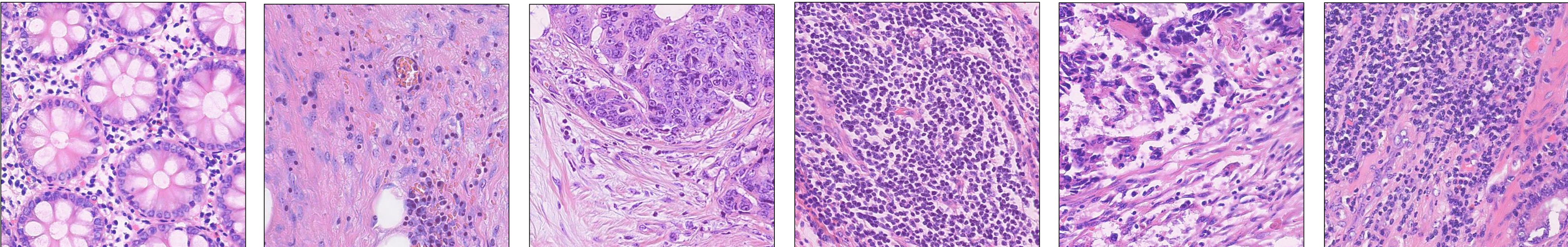


Figure 4: Target image having the color and intensity distributions which should be matched.

Figure 5: H&E images before and after applying stain normalization.