

EVALUATING THE IMPACT OF COLOR NORMALIZATION ON KIDNEY IMAGE SEGMENTATION

Sai Javvadi

University of Louisville, Louisville Kentucky, USA

ABSTRACT

The role of deep learning in the recognition of morphological structures in histopathological data has progressed significantly. But, less intensive preprocessing stages and their contribution to deep learning pipelines is often overlooked. Color normalization (CN) algorithms are among the most prominent methods in this stage, and they work by standardizing the staining pattern of a dataset. However, the impact of various color normalization algorithms on the detection of glomeruli functional tissue units (FTUs) in kidney tissue data has not been explored before. An advanced deep learning architecture was built with the U-NET segmentation model. The U-NET model is an architecture that specializes in the segmentation of biomedical data. A dataset of 15 kidney whole slide images (WSIs), each annotated with locations of glomeruli FTUs were processed and subsequently normalized according to three 3 different conventional color normalization techniques (Reinhard, Vahadane, Macenko), and fed into a U-NET model. The dice score coefficient (DSC) was used to compare the results of each run. It was determined that color normalization algorithms significantly impact the segmentation results of deep learning algorithms, with the Reinhard algorithm being the best technique. The implications of this work are immense, as it could contribute to the proliferation of color normalization techniques in preprocessing deep learning workflows, which could improve general segmentation accuracies.

KEYWORDS

Deep Learning, Color Normalization, Histopathology, Kidney, Glomeruli

1. INTRODUCTION

1.1. Overview

The application of deep learning within the context of medical imaging has allowed for faster and more accurate analysis compared to conventional methods of analysis that rely on a pathologist. The advantages that deep learning provides in this context can be highlighted when dealing with organ tissues that exhibit great heterogeneity, such as the kidney. The great diversity of tissue within the kidney makes it especially difficult for pathologists to annotate, consequently making deep learning techniques focused on the annotation of kidney tissue more valuable and significant [1]. Thus, exploring ways to make deep learning models trained on histopathological kidney tissue data more accurate and useable within a clinical setting is imperative. By being able to effectively evaluate particular techniques that could contribute to the enhancement of deep learning model performance and applicability when applied to kidney tissue data, similar strides could be taken with other kinds of tissue data. The widespread application of deep learning in pathological contexts, and thus for the diagnosis and prognosis of multi-organ diseases, could revolutionize the medical sphere [2].

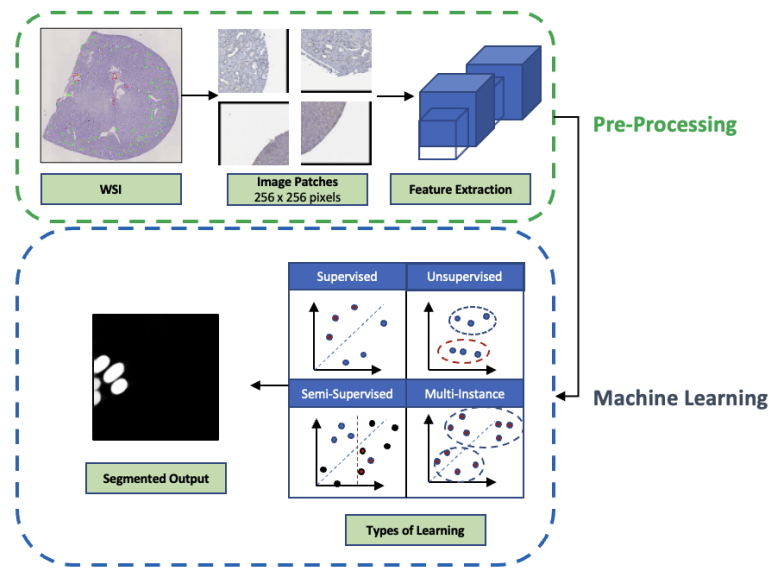


Figure 1. Histopathological Image Segmentation Workflow

The purpose of this research is to evaluate the impact of various, conventionalized color normalization techniques on a self-engineered U-NET deep learning model focused on analyzing kidney tissue data. This would allow for the changes in model performance according to color normalization technique to be quantified, and could allow for the greater application of particular color normalization techniques in more contexts. This could also effectively improve model performance and make strides in achieving clinical integration.

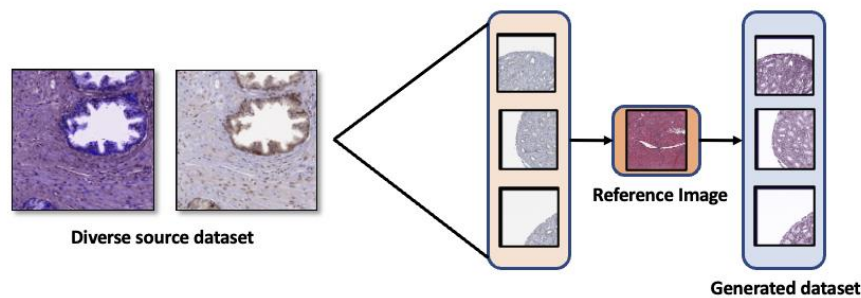


Figure 2. Color Normalization Workflow – Pattern of reference image applied to diverse source dataset for normalized generated dataset

1.2. Literature Review

Previously developed deep learning models trained to identify glomerular functional tissue units have often emphasized the performance of the U-NET model for its high accuracy [3, 4]. The U-NET model was proposed in 2015, and is a deep learning model especially compatible with biomedical imaging due to its unique architecture [5]. Proposed improvements in previously referenced works are the inclusion of a preprocessing technique that normalized stains in the datasets [3]. Further work involves the development of a computer-aided diagnostic (CAD) model that is capable of not only segmenting, but also classifying glomeruli [6]. Although these works demonstrate a variation in methodology in terms of the model leveraged (an ANN rather

than the more commonly used CNN), an emphasis on the experimentation of various preprocessing techniques, namely color normalization, is generally absent. Previous work that has focused on the application of various color normalization algorithms and their impact on deep learning models trained on different kinds of histopathological data [6] allows for the techniques that are potentially effective with kidney tissue data to be narrowed down, despite this work not experimenting with this type of tissue in particular. Other work that focused on gauging the impact of color normalization techniques on other kinds of biomedical data have found that it successfully tamed variability in data [7].

The lack of research surrounding the application and evaluation of color normalization techniques on kidney histopathological images is a clear research gap. This project aims to address this gap by testing 3 color normalization techniques that are most commonly used in order to determine the impact of color normalization on deep learning model performance, potentially allowing for heightened model performance and greater clinical applicability. This project hypothesizes that every single normalization technique will aid in a model that improves upon the performance of a deep learning model trained solely on the original, unnormalized data. Furthermore, of the 3 color normalization algorithms being tested – Vahadane, Reinhard, and Macenko – it is hypothesized that the Reinhard algorithm will perform the best due to its frequent use in literature.

1.3. U-NET Model Overview

The U-NET model is able to generally provide better segmentation accuracies when applied to histopathological data due to its particular architecture, as seen in Figure 6. The first half of its U-shape is known as the contracting path, which primarily consists of convolutional and max pooling layers. These layers are used to extract features, and to reduce the dimensionality of feature maps. The use of these layers allowing for complex and high-level features to be captured. The second half of the model is composed of the expanding path. This path creates the segmentation map and leverages transposed convolution (or Up-convolution) to increase the resolution of the feature maps. Essentially, this path transforms the compressed representation from the contracting path to a segmentation map in the original resolution. One of the most unique aspects of the U-NET architecture are the skip connection layers (lines connecting the contracting path to the expanding path). These connections allow for the preservation of very specific spatial information. This is especially beneficial when applied to extremely complex and detailed data like that of histopathological data. Consequently, when applied to this kind of data, more accurate and precise segmentations can be achieved.

2. METHODOLOGY

Kidney tissue data is derived from the “HuBMAP: Hacking the Kidney Dataset”. This dataset presents 20 total Formalin Fixed Paraffin Embedded (FFPE) Periodic Acid-Schiff (PAS)-stained kidney whole slide images (WSIs). The data is pre-annotated with the locations of Glomeruli Functional Tissue Units (FTUs). From each image, individual tiles of resolution 512 x 512 are obtained; this processing stage results in roughly 2600 images. In order to proliferate the data, augmentative procedures are used: random cropping, random mirroring, random jittering, and random noise. Examples of these augmentation on data is demonstrated in Figure 4. These augmentative procedures create transformation within the data so as to create more variations and to expand the size of the dataset. Furthermore, augmentations contribute to a more generalized and accurate deep learning model. This proliferated dataset is then portioned for training/testing purposes with a 90/10 ratio (90% of the dataset is allocated for training, while the other 10% is set aside for testing).

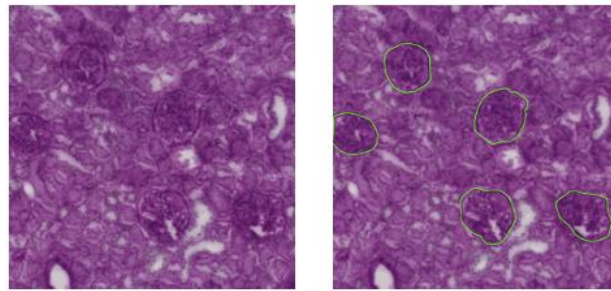


Figure 3. Example of Kidney Data - Tile (left) and Annotated Tile (right)

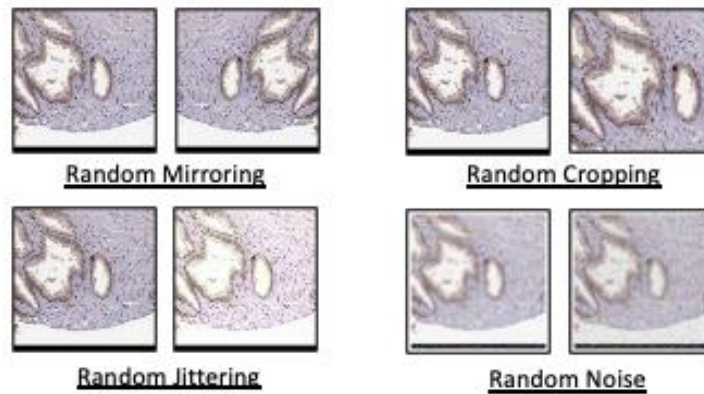


Figure 4. Augmentative Functions - Left: Original Image; Right: Augmented Image

A commonly selected reference image is chosen based upon the effectiveness of the reference image established by successful teams who had previously worked with the dataset. The original dataset is normalized based upon a singular reference image by each of the three-color normalization algorithms. In total, there are 4 datasets (the original, and 3 normalized). An array of color metrics is applied to each dataset to gauge the impact of color normalization on the quality of the dataset. Subsequently, each dataset is fed into a U-NET deep learning model trained to recognize glomeruli FTUs. The dice score coefficient (DSC) is used to evaluate the performance of the model for each dataset, effectively comparing the impact of color normalization on model accuracy. These results leverage 10-fold cross validation to ensure statistical significance.

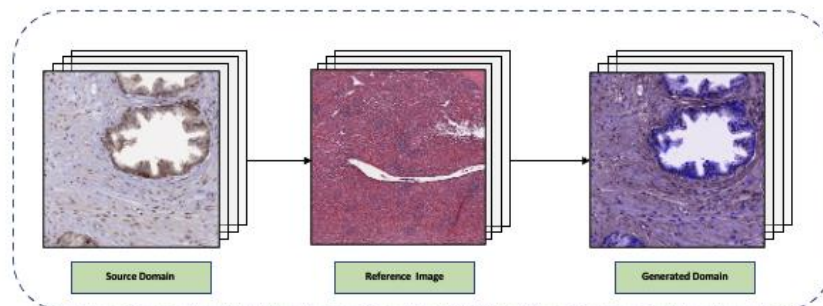


Figure 5. Applying Reference Pattern to Source Dataset – A generated dataset more aligned in color is generated by applying the pattern of the reference image to the original dataset

2.1. Color Normalization

Color Normalization (CN) algorithms were specifically chosen for experimentation due to their prior success in other tissue-based segmentation algorithms [8]. The successful use of the Reinhard, Vahadane, and Macenko CN algorithms for medical segmentation tasks may be potentially translated to glomeruli detection algorithms.

Reinhard et al developed a globalized technique for color normalization, in which the source image receives the mean color values of the target, or reference image. This results in a generated image that is very similar to that of the reference [9].

Macenko and Vahadane developed color normalization techniques that leveraged stain separation. These separated stain maps are normalized and combined in a fashion that relies on the stain color of the reference image [10].

2.2. Architecture

A constant architecture must be used because with this constant, an accurate measure of the impact of CN algorithms would not be available. Furthermore, studies that test various data augmentations keep their models constant as changing the models can impact the overall accuracy, interfering with the comparison of different preprocessing algorithms and creating distorted comparisons. For example, previous works that evaluate the strength of color normalization methods have kept particular model architectures as constants in their experiments [20]. In this particular project, the U-NET model architecture is specifically used is due to the prevalence of its usage in various medical segmentation purposes [11]. The architecture is specialized to work with medical data that is meant to be segmented, and operates in a fast and precise manner [12]. In the context of identifying glomeruli specifically, recent literature has widely used the U-NET algorithm [13]. Therefore, using U-NET as a constant architecture ensures a good baseline performance while practically measuring any added benefit of a added CN algorithm.

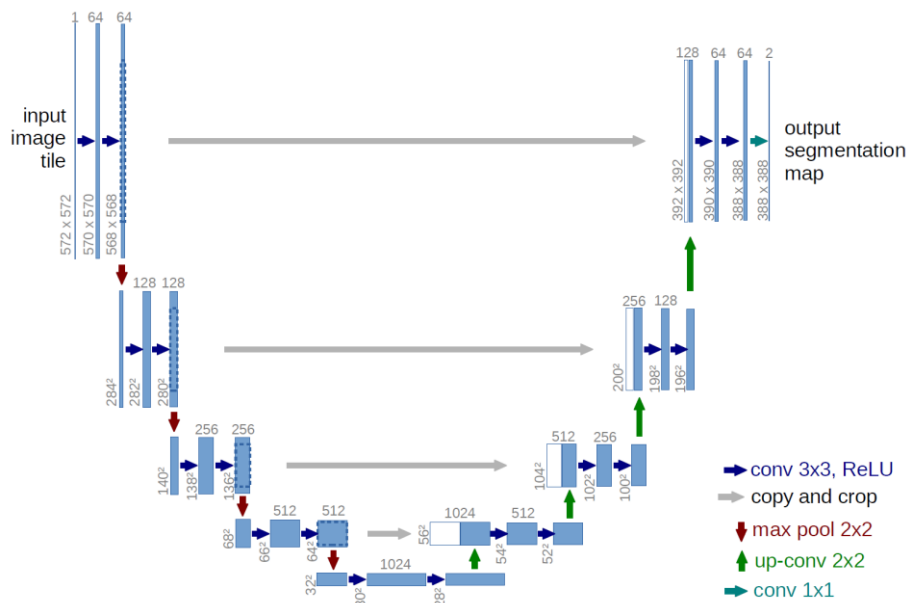


Figure 6. U-NET Deep Learning Model Architecture – Retrieved from [14]

The particular U-NET model architecture relies on the Dice Score Loss function, which quantifies the difference in segmentation between a model's predicted segmentation and the ground truth. Furthermore, the Adam optimizer is leveraged at learning rate $1e-4$ in order to adjust the weights of the deep learning model so as to achieve a set of weights capable of optimal model performance. Furthermore, the model is trained on 100 epochs with a batch size of 8. This means that the model goes through the entirety of the training data 100 times, and is updated whenever it processed 8 images.

To reiterate, the novelty of this work is derived from the direct comparison of model performance when trained on datasets processed with various color normalization techniques (Reinhard, Vahadane, & Macenko). In the particular context of segmenting Glomeruli FTUs in kidney images, there is a lack of this work. Demonstrating the effectiveness of these processing techniques with morphologically complex data may contribute to its implementation in wider scales, and result in residual benefits such as mitigating bias in deep learning models (bias accumulates from heavy diversity in stains).

2.3. Performance Metrics

Once the deep learning model is developed, performance metrics need to be compared to determine whether models with CN preprocessing stages present any added benefits. The usage of performance metrics has long characterized the effectiveness of machine learning algorithms [15]. Therefore, using the correct performance metrics is vital. For segmentation, prevalent performance metrics include the Dice Score Coefficient (DSC), IoU (Intersection-over-Union), and accuracy. Previous works have used the DSC and IoU metrics to directly compare and evaluate various models trained of kidney-based glomerular data [16]. Consequently, the use of these metrics as a ground for comparison among a variety of models suggests that it is a reputable method for the accurate reflection of model performance. For this research, the DSC was adopted as a means to quantify model performance.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

Equation 1. Dice Coefficient

The DSC value measures the intersection between two areas. It is $2 * (\text{area of intersection}) / \text{total number of pixels in two areas}$. The greater this value, the more indicative of a model's strength, as the predicted segmentation is closer to the ground truth (real) segmentation.

The use of PAS-stained data to feed into U-NET algorithms is commonplace in this field. Therefore, comparing this workflow to one that uses a color normalization algorithm will provide valuable insight into the latter's impact on model performance while also being practical. This methodology will allow the impact of color normalization algorithms to properly be evaluated and potentially make large strides in the use of CN algorithms as a general preprocessing implementation. There were two primary phases for testing the impact of color normalization with glomeruli and histopathological data. The first step was to choose three state-of-the-art color normalization methods: Reinhard, Macenko, and Vahadane. In order to compare the 3 different methods of color normalization tested (the 3 experimental variables being the color normalization techniques being tested, and the control being the baseline workflow without normalization), it was imperative that an array of color metrics, or measurements that indicate particular qualities of

an image’s color nature, was used. Rather than use a singular color metric to compare the performances of these different methodologies, an array of metrics was adopted due to the fact that since color can be measured in a myriad of ways: intensity, hue, saturation, etc. The use of a singular metric would not be indicative of the overall performance of a normalization technique [1]. The particular metrics adopted were based on previous work that evaluated the impact of color normalization through these metrics [17].

The particular metrics adopted include FSSIM, or the functional similarity index measure which screens for the quality of a new image in relation to an original; UQI, or universal quality index, which reflects the overall quality of an image; PCC, or the Pearson Correlation Coefficient, which measures the strength of the relationship between two images.

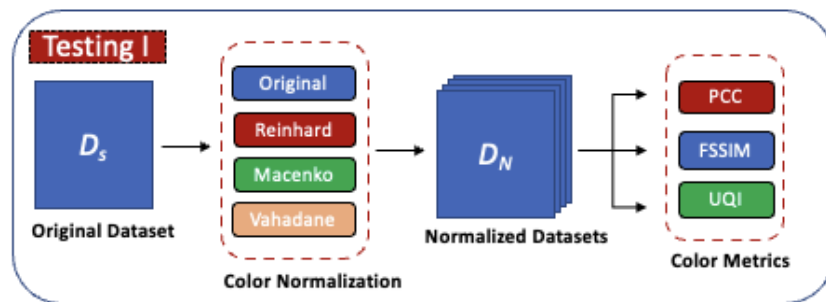


Figure 7. First Testing Phase Workflow – The original dataset is normalized with each CN algorithm. 3 different color metrics (PCC, FSSIM, and UQI) are subsequently applied to each dataset in order to gauge for quality.

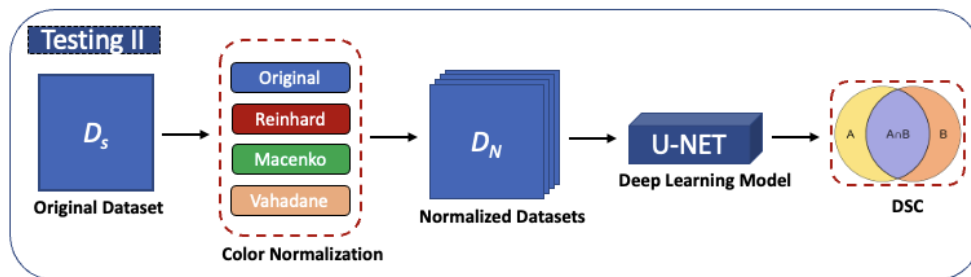


Figure 8. Second Testing Phase Workflow – The normalized datasets, along with the original, are subsequently fed into a deep learning model. Model performance is evaluated with the DSC metric.

2.4. Model Training Environment

With this methodology, various color normalization algorithms can be properly compared in multiple facets. Through the application of different color metrics, the quality of the data they produce can be evaluated. More importantly, the development of a U-NET deep learning model allows for the implications of the differently normalized datasets on model performance to be properly evaluated. The development of the aforementioned architectures will be conducted natively using the Google Collaboratory service. By utilizing cloud-based high-end GPUs, the training and testing requirements of this project may be fulfilled significantly faster.

3. RESULTS

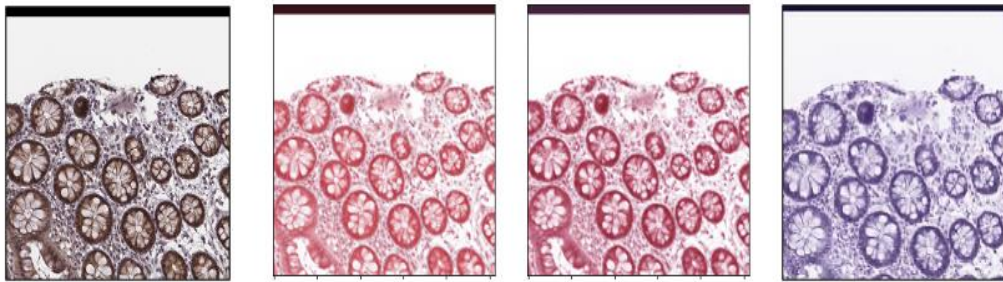


Figure 9. Example of Normalized Images – From left to right: Original Image, Macenko Normalized, Vahadane Normalized, Reinhard Normalized

3.1. Color Metric Performance Results

Comparing the qualities of the color normalized datasets in Figure 10, Vahadane possesses the composite highest quality. While the Reinhard-normalized dataset possesses the highest average PCC score, the Vahadane-normalized data boasts the highest average score for the FSSIM and UQI metrics. Since both the FSSIM and PCC metrics evaluate the quality of an image in relation to another, these metrics offer conflicting narratives about the quality of the generated images in relation in to the original. However, the UQI metric analyzes general images quality and does not do so in relation to another image. Based upon these results, a general ranking of the quality of each of the normalized datasets can be assigned (in highest to lowest order): Vahadane, Reinhard, Macenko.

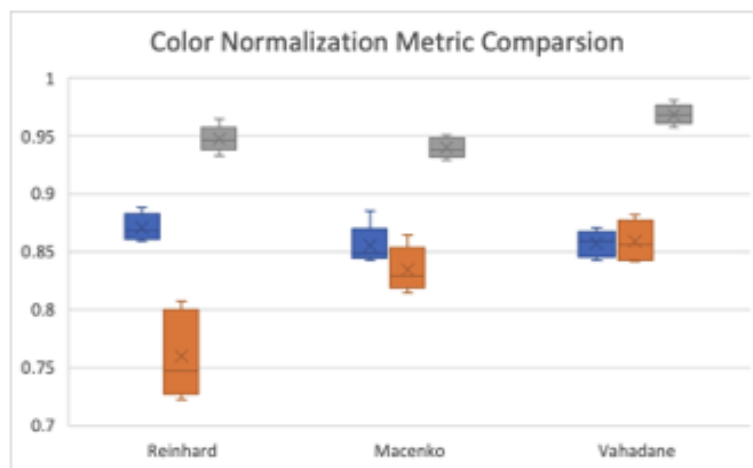


Figure 10. Color Metric Array Comparison - Blue: PCC; Orange: FSSIM; Gray: UQI

From this initial analysis of the normalized data, it was deduced that since the Vahadane-normalized data reflected attributes that suggested it was of higher quality than the other normalized datasets, that there would potentially be a correlation to better performance. This is because higher quality data may be more meaningful and beneficial for feature learning, contributing to better model performances.

The quality and effect of normalization on data can be further visualized by comparing the nature of the normalized data to that of the un-normalized, original dataset. The images were compared by using the average Hue and Saturation channel ratios of every single image in the dataset. This was done by converting each RGB image in all datasets into the HSV (Hue Saturation Value) file format. Then, the hue and saturation channels were extracted and compared for each image. This method is reflective of the effectiveness of color normalization, particularly within histopathological images with a bimodal nature, or of a nature in which two hues dominate the image when visualized with the hue channel [18].

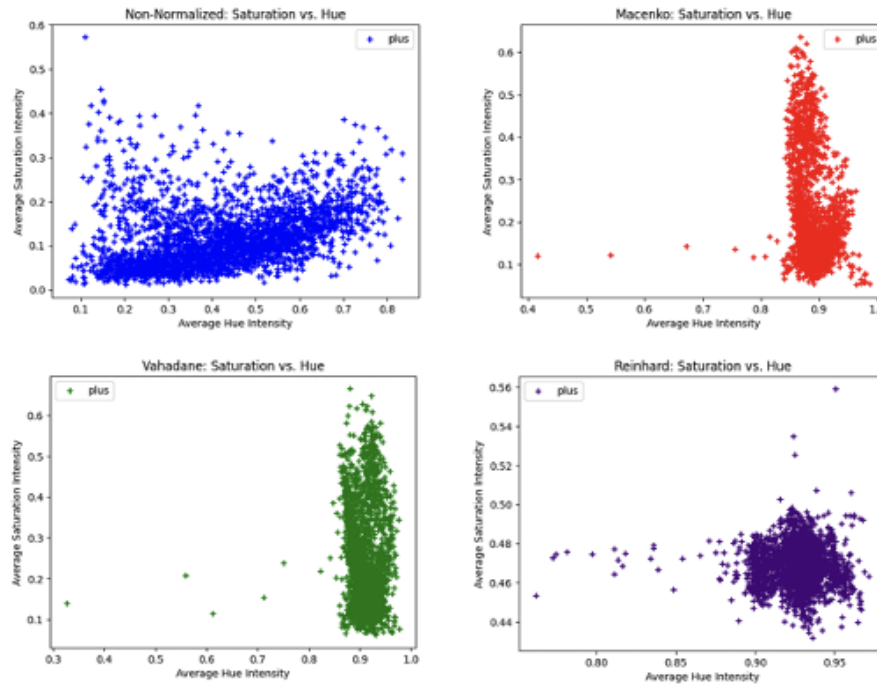


Figure 11. Average Hue vs. Saturation Intensities for Non-Normalized and Normalized Datasets

As visualized above, every single normalization technique was able to effectively normalize the original data. The average Saturation vs. Hue intensity is significantly more concentrated for each of the normalized data plots than that of the non-normalized data, which is indicative of effective color normalization. Furthermore, the styles and natures of the normalization techniques can also be observed with these plots. The Macenko and Vahadane plots are very similar, with a cluster of points congregated in a vertically linear fashion towards the right-most portions of the plot. In contrast, the Reinhard plot depicts a cluster of points arranged in a circular fashion, still closer to the right-edge of the plot. This is expected since the Macenko and Vahadane techniques operate very similarly, while the Reinhard method takes a very different approach in color transformation.

Analyzing the plots of the normalized datasets, the quantity of outliers can be observed. Both the Vahadane and Macenko normalized datasets seem to have a limited numbers of points that are distant from a central cluster. In contrast, the Reinhard normalized dataset seems to have a significant number of outlier points. Many points on the Reinhard plot are very distant from the central cluster. The degree of alignment, although strong with all normalized datasets, appears to be weaker in the Reinhard normalized dataset. This could potentially be indicative of less effective normalization.

3.2. Model Segmentation Results

It is critical that the effects of these color normalization techniques on the segmentation of the histopathological dataset be effectively tested. Since color normalization effects the nature of an image, and deep learning models segment these images to precisely identify particular morphological features, the effect of these techniques on deep learning segmentation was critical for a thorough understanding of the practicality that color normalization offers in the clinical context.

Therefore, it was important to test for the effectiveness of color normalization techniques on the identification of glomerular structures within kidney WSIs.

For the statistical validation of this data K-folds cross validation (CV) was used; K-folds CV is a commonly used tool to ensure that the results of machine learning performances are not merely coincidental [19]. As demonstrated in Figure 10, the testing partition of the data is varied with each run. With 10-fold CV, each dataset (original, and 3 normalized) is fed into the model 10 times, each time with a different validation set. Each run returned a value for the DSC, which reflects a generalized performance of the model’s predicted segmentation accuracy.

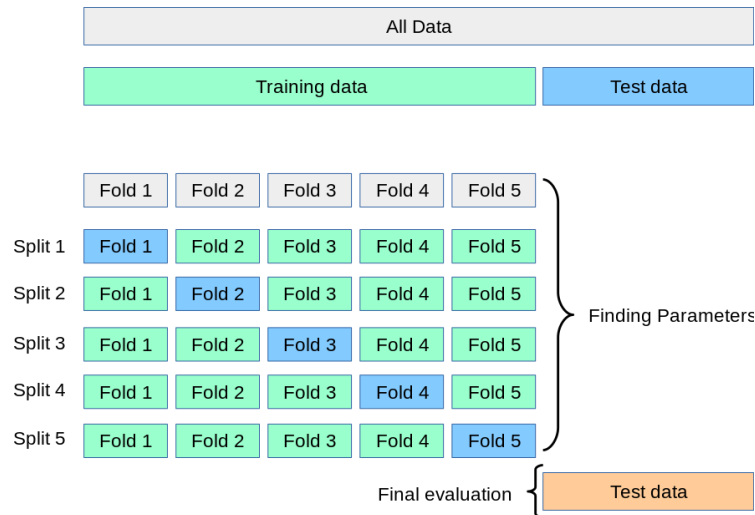


Figure 12. K-Fold Cross Validation – Retrieved from [20]

Table 1. Average DSC Performances – Averages the DSC score outputted from a model trained on each dataset

Average DSC with 10-Fold CV				
Dataset	Original	Macenko	Vahadane	Reinhard
HuBMAP	0.7963	0.8712	0.8124	0.8842

As observed in Table 1, the Reinhard algorithm boasts the highest average DSC value. As this result is the culmination of 10 different runs, each with different folds of training and test data, this result is treated as non-coincidental and statistically-validated.

These results suggest that although every single normalization technique improved upon the baseline performance of the deep learning model, the Reinhard algorithm was most successful in doing so. The Reinhard algorithm was able to achieve an average DSC of 0.8842, a massive leap

in performance compared to the average DSC of 0.7963 with the original dataset. Furthermore, the color metrics used to test the general quality of the produced images indicated that the Vahadane technique produced better quality images, but despite this, did not outperform the Reinhard technique. This result reinforces the prominence of the Reinhard algorithm in deep learning workflows that do utilize color normalization.

4. CONCLUSION

To conclude, it was found that color normalization has a significant impact on the identification of glomeruli within PAS-stained kidney whole slide images. Color normalization involves standardizing stain variations that may exist within a dataset. Of the 3 common color normalization algorithms tested, it was determined that all three of them resulted in higher segmentation performance for the tested dataset, as according to the Dice Score Coefficient, than the original dataset. This means that when fed with the normalized data, deep learning models, regardless of what normalized data it was fed, was able to recognize more glomeruli with higher precision than what it was able to when fed the original, non-normalized data. The quality of the segmentations produced also differed drastically. Often times, the segmentation produced by the model that was fed some type of normalized data was significantly more precise in a variety of aspects. Greater precision can be exhibited by the sheer number of structures within an image that was able to be identified by a deep learning model, and the edges of the segmentations (precise segmentations have rougher edges while less precise segmentation are rounder and more connected together, representing less precision). In order to obtain these results, it was imperative that the deep learning model used for experimentation was kept constant. Interestingly enough, it was determined that the original data was generally of a higher quality according to a myriad of color metrics applied. This supports the idea that quality does not translate to segmentation performance.

The implications of this research are strong – as more attention is paid to the application of certain CN algorithms on the segmentation of not only kidney tissue, but other types of histopathological images, the biases that are often inevitable in histopathological datasets may be combatted. These are the same biases that render many, even highly-accurate models, impractical within a hospital setting. This research hopes to support a narrative, that, color normalization can aid in the reduction of bias present in stain data and make deep learning algorithms more suitable and safer to use in a clinical setting.

4.1. Limitations

This research is limited by only sourcing data from a singular dataset. By agglomerating data from diverse sources, the results of this study and corresponding application can be reinforced. Furthermore, the size of the dataset itself was quite small (~2600 images without augmentative procedures). In future studies, similar data from other datasets can be sourced and combined for larger datasets. Furthermore, this study is limited by the use of only a singular deep learning model (U-NET model) being used to test the effectiveness of normalization. Since only a single deep learning model is used, the impact of different normalization techniques is not fully portrayed. In future studies, testing multiple architectures will be considered so that more representative results can be conveyed.

4.2. Future Direction

For future research, the application of new color normalization techniques can be explored and compared to the effectiveness of non-normalized, original data and even compared against conventional, state-of-the-art color normalization techniques. In this work, conventional techniques were tested. However, with the breakthroughs of new generative models, such as generative adversarial networks (GANs) in the field, new color normalization techniques are being developed which could have significant implications for the landscape. Furthermore, more model architectures other than U-NET can be tested.

REFERENCES

- [1] Huo, Y., Deng, R., Liu, Q., Fogo, A. B., & Yang, H. (2021). AI applications in renal pathology. *Kidney International*, 99(6), 1309–1320. <https://doi.org/10.1016/j.kint.2021.01.015>
- [2] Furlow, B. (2017). Deep learning poised to revolutionise diagnostic imaging. *The Lancet Respiratory Medicine*, 5(10), 779. [https://doi.org/10.1016/s2213-2600\(17\)30292-8](https://doi.org/10.1016/s2213-2600(17)30292-8)
- [3] Shubham, Shubham, et al. “Identify Glomeruli in Human Kidney Tissue Images Using a Deep Learning Approach.” *Soft Computing*, vol. 27, no. 5, 23 Aug. 2021, pp. 2705–2716, <https://doi.org/10.1007/s00500-021-06143-z>.
- [4] Wilbur, D. C., Smith, M. L., Cornell, L. D., Andryushin, A., & Pettus, J. R. (2021). Automated identification of glomeruli and synchronized review of special stains in renal biopsies by Machine Learning and Slide Registration: A cross-institutional study. *Histopathology*, 79(4), 499–508. <https://doi.org/10.1111/his.14376>
- [5] Piórkowski, A., & Gertych, A. (2018). Color normalization approach to adjust nuclei segmentation in images of hematoxylin and eosin stained tissue. *Advances in Intelligent Systems and Computing*, 393–406. https://doi.org/10.1007/978-3-319-91211-0_35
- [6] Cascarano, Giacomo Donato, et al. “A Neural Network for Glomerulus Classification Based on Histological Images of Kidney Biopsy.” *BMC Medical Informatics and Decision Making*, vol. 21, no. S1, Apr. 2021, <https://doi.org/10.1186/s12911-021-01650->
- [7] Otálora, S., Atzori, M., Andrearczyk, V., Khan, A., & Müller, H. (2019). Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology. *Frontiers in Bioengineering and Biotechnology*, 7. <https://doi.org/10.3389/fbioe.2019.00198>
- [8] Pontalba, J. T., Gwynne-Timothy, T., David, E., Jakate, K., Androutsos, D., & Khademi, A. (2019). Assessing the impact of color normalization in convolutional neural network-based nuclei segmentation frameworks. *Frontiers in Bioengineering and Biotechnology*, 7. <https://doi.org/10.3389/fbioe.2019.00300>
- [9] Lakshmanan, B., et al. “Stain Removal through Color Normalization of Haematoxylin and Eosin Images: A Review.” *Journal of Physics: Conference Series*, vol. 1362, no. 1, 1 Nov. 2019, p. 012108, <https://doi.org/10.1088/1742-6596/1362/1/012108>.
- [10] Kang, Hongtao, et al. “Stainnet: A Fast and Robust Stain Normalization Network.” *Frontiers in Medicine*, vol. 8, 5 Nov. 2021, <https://doi.org/10.3389/fmed.2021.746307>.
- [11] Agraz, J. L., Grenko, C. M., Chen, A. A., Viaene, A. N., Nasrallah, M. D., Pati, S., Kurc, T., Saltz, J., Feldman, M. D., Akbari, H., Sharma, P., Shinohara, R. T., & Bakas, S. (2022). Robust image population based stain color normalization: How many reference slides are enough? *IEEE Open Journal of Engineering in Medicine and Biology*, 3, 218–226. <https://doi.org/10.1109/ojemb.2023.3234443>
- [12] Ronneberger, O. (2017). Invited talk: U-net convolutional networks for biomedical image segmentation. *Informatik Aktuell*, 3–3. https://doi.org/10.1007/978-3-662-54345-0_3
- [13] Godwin, L. L., Ju, Y., Sood, N., Jain, Y., Quardokus, E. M., Bueckle, A., Longacre, T., Horning, A., Lin, Y., Esplin, E. D., Hickey, J. W., Snyder, M. P., Patterson, N. H., Spraggins, J. M., & Börner, K. (2021). Robust and generalizable segmentation of human functional tissue units. <https://doi.org/10.1101/2021.11.09.467810>
- [14] Konovalenko, Ihor, et al. “Research of U-Net-Based CNN Architectures for Metal Surface Defect Detection.” *Machines*, vol. 10, no. 5, 29 Apr. 2022, p. 327, <https://doi.org/10.3390/machines10050327>.

- [15] Botchkarev, A. (2019). A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14, 045–076. <https://doi.org/10.28945/4184>
- [16] Li, X., Davis, R. C., Xu, Y., Wang, Z., Souma, N., Sotolongo, G., Bell, J., Ellis, M., Howell, D., Shen, X., Lafata, K. J., & Barisoni, L. (2021). Deep learning segmentation of glomeruli on kidney donor frozen sections. *Journal of Medical Imaging*, 8(06). <https://doi.org/10.1117/1.jmi.8.6.067501>
- [17] Roy, S., kumar Jain, A., Lal, S., & Kini, J. (2018). A study about color normalization methods for histopathology images. *Micron*, 114, 42–61. <https://doi.org/10.1016/j.micron.2018.07.005>
- [18] Tellez D;Litjens G;Báñdi P;Bulten W;Bokhorst JM;Ciompi F;van der Laak J; “Quantifying the Effects of Data Augmentation and Stain Color Normalization in Convolutional Neural Networks for Computational Pathology.” *Medical Image Analysis*, pubmed.ncbi.nlm.nih.gov/31466046/. Accessed 31 July 2023.
- [19] Wang, C.-M., Huang, Y.-H., & Huang, M.-L. (2006). An effective algorithm for image sequence color transfer. *Mathematical and Computer Modeling*, 44(7-8), 608–627. <https://doi.org/10.1016/j.mcm.2006.01.029>
- [20] “3.1. Cross-Validation: Evaluating Estimator Performance.” *Scikit*, scikit-learn.org/stable/modules/cross_validation.html. Accessed 5 Aug. 2023.

AUTHOR

Sai is a researcher associated with the university of Louisville's Biomedical Imaging Lab. His concentrated intrest is exploring and applying new deep learning techniques to biomedical data.

