

Investigating the Role of Colour in Deep Neural Network -Based Nuclei Segmentation

Justin T. Pontalba

Dept. of Electrical, Computer, and Biomedical Engineering

Ryerson University

Toronto, Canada

j pontalb@ryerson.ca

Abstract—A pre-processing method prior to segmentation, called colour normalization, has traditionally been applied to image data as an attempt to reduce variability and thus improve algorithm performance. While colour normalization demonstrated benefit to traditional classification and segmentation methods, it was not until recently where colour normalization was demonstrated to negatively impacted segmentation performance on various convolutional neural network architectures (CNN). The proposed work aims to further investigate the role of colour in CNNs by quantifying and comparing the response of a CNN model to red, green, blue (RGB), and grayscale images at various layers using Jensen-Shannon Divergence. In addition, at the same layers, we produce the reconstruction of an image given a predicted encoding in order to examine if colour features were important to predict such encoding. By analyzing the response of the CNN models to colour presence and subtle colour changes, and by studying the reconstructions at various layers, a better understanding of CNNs can be realized and nuclei segmentation pipelines can be better optimized.

Index Terms—colour normalization, computational pathology, convolutional neural networks, deep learning, nuclei segmentation

I. INTRODUCTION

Cancer is the number one leading cause of death in Canada, and by the end of 2019, an approximated 220 400 Canadians will be diagnosed with cancer. Methods for pre-screening and routine analysis of clinical data have been used for early intervention and to identify risk factors [1]. Initial diagnosis and management of cancer is achieved through physical examination, laboratory tests, imaging tests, and biopsies. Depending on the type and location of the suspected lesion, biopsies are less invasive and reveal characteristics of the tissue. Biopsy analysis is completed by a pathologist and examines nuclei features such as morphology and spread. These features describe the shape, colour, and spatial organization of nuclei respectively. How abnormal tissue appears according to these characteristics can be quantified through a proxy metric called histological grade.

Histological grading is common among many cancers and begins with the staining of tissue slides using hematoxylin and

eosin (H&E). H&E stains are applied to tissues to increase contrast between nuclei and stromal structures. For instance, hematoxylin, which appears as purple, binds to nucleic acids in the nuclei. Similarly, eosin, appearing pink, binds to the cytoplasm of cells [2]. It is under these stain combinations where histological grade is obtained. Unfortunately, histological grading is time consuming and the interpretation is error-prone due to inter-rater variability [3] [4]. With the onset of digital whole-slide scanners, automated algorithms for pathological analysis have been proposed as a solution to reduce subjectivity and variability and improve clinical workflow.

A. Related work

Early works for automated pathological analysis focused on nuclei detection. However, the potential for richer information is greater when the entire nuclei is captured. Features such as colour, texture, and size can be extracted from individual nuclei, and various tissue regions can be analyzed individually or as a whole. As such, later works propose automated nuclei segmentation pipelines.

Automated segmentation of nuclei is an important step prior to automated nuclear grading. However, automating this task is difficult due to the amount of variability contained in digital pathological data. Disease morphology, patient anatomy, and vendor variation (between digital scanners and stains) are barriers to developing generalizable automated systems. This is especially apparent for traditional frameworks when nuclei are clustered and overlapping [5] [6]. Morphological processing [7], hand-crafted feature design and classification [8], unsupervised clustering [9], and supervised approaches that classify each pixel into different categories: nuclei or background [10] [11] [12] are some examples of traditional segmentation frameworks. In all cases, data variability is the main barrier that prevents the development of robust and generalizable algorithms. Over the last few years, deep convolutional neural networks (DCNN), have become popular in the analysis of digital tissue specimens due to their dominating performance on large and variable datasets [13]. Computational pathology has leveraged the usage of DCNNs for these reasons. Very

Funding received for this project is from the Mitacs Accelerate Grant (IT12249).

early DCNNs, such as LeNet, AlexNet, GoogleNet, and VGGNet, modeled their architectures as a series of successive convolutional, pooling, and fully connected layers. While effective, variations of these models were computationally expensive due to the fully connected layers contained in the architectures. To combat the issue caused by fully connected layers, fully convolutional networks (FCNs) were developed. FCNs do not contain any densely or fully connected layers, utilize skip connections, and produce semantic predictions of the same size as the input. Most notably, the U-Net architecture has been widely used and adapted due to quick prototyping and segmentation accuracy. U-Net was developed by connecting features from the downward path with up-sampled outputs at various layers. By connecting these paths, high resolution features can be localized at the output layers.

Regardless of successful DCNN architectures such as U-Net, data variability continues to be a barrier to generalization especially in digital pathology applications. A solution to reduce variability and improve generalization is a pre-processing step called colour normalization. Colour normalization aims to reduce the stain or colour variability contained in digital pathology data by transforming datasets into a common space. Success of colour normalization was most prominent in traditional segmentation frameworks, as well as, DCNNs for classification tasks [14] [15] [16] [12]. However, in recent works, colour normalization did not significantly improve applications in segmentation. This is demonstrated in works such as [17] and [18]. Both works observed the impact of normalization on segmentation performance. Lafarge et al demonstrated that rather than normalization, stain and colour augmentation increased segmentation performance, and [18] demonstrated that un-normalized-based models outperformed any colour normalized-based models. [18] also stipulated that colour normalization can introduce image artefacts that negatively impact model performance especially when a dataset not seen during training is used. In either work, while the segmentation validation metrics were examined, colour perception and representation in DCNNs were not further investigated or quantified. This is important because how colour is perceived by DCNNs is suspected to be directly related to performance, since colour is the feature that is being altered in each paper. Feature perception and representation in DCNNs is an important concept as it increases interpretability of DCNN based algorithms and systems. This is especially important in health and medicine applications, as interpretability is important in making decisions related to patient care from both a safety and policy perspective.

A detailed review by [19] examines several research methods that attempt to understand deep neural-network representations. Visualizing intermediate layers, diagnosing DCNN representations, disentangling patterns encoded in filters, and building explainable models are examples of how researchers aim to increase interpretability of CNNs to move towards explainable artificial intelligence.

In digital pathology applications, interpretability methods align with the visualization of intermediate layers to analyze

how filters respond to varying inputs. For instance, Hsu et al investigated how activation features work as morphological descriptors for specific cells or tissue [20]. In another case, Korbar et al analyzes activation features that correspond to different classes or colorectal polyps [21]. The activation features of sub-patches are projected back to the original image to understand regions or features that are responsible for the classification output. Visualizing the response of various layers in this way increases the potential to understand how DCNNs respond to subtle changes in colour.

B. Contributions in this paper

The work presented in this paper emphasizes studying how changes in colour can affect the response of filters at various layers of a DCNN. The proposed paper extends the work of [18] by investigating the importance of colour as a feature. [18] concluded that colour normalization impacts segmentation performance on data that went unseen during training but does not largely impact performance on seen data. Our proposed approach both quantitatively and qualitatively examines the response of DCNN layers to changes in colour resulting from colour normalization. Inspired by a previous work [22], the Jensen-Shannon divergence metric is used to quantify how sensitive the model is to changes in colour at different layers. Furthermore, the filters which demonstrate the highest average activation value are compared and analyzed. Lastly, filters are inverted at the same layers of the previous experiment to analyze if colour is an important feature. If original colour is contained in the inverted image, it is suspected that colour is recognized as an important feature to the DCNN. However, if colour is not present then colour may not be as important.

The paper is organized as follows. In section II materials and methods are detailed. Section III highlights results of the experiments and Section IV outlines discussions and conclusions.

II. MATERIALS AND METHODS

A. Data

The image patches required to test and evaluate our methods were originally obtained by [14] from The Cancer Genome Atlas (TCGA). The dataset contains hematoxylin and eosin (H&E) stained patches with manually annotated nuclei. Nuclei annotations were created under the supervision of a pathologist by undergraduate students. The whole slide images in which the patches were cropped from were digitized under 40X magnification. In total, 29 1000 x 1000 region of interest images were cropped from the whole slide, and the dataset is comprised of breast, lung, kidney prostate, bladder, colon, and stomach tissue types. The tissue types contain benign and malignant diseases and exhibit highly variable staining, intensities, colours, and nuclei appearances across organs. Such characteristics makes the dataset suitable for the application of colour normalization and the proposed evaluation.

B. Experimental Design

Figure 1 depicts the experimental design used to address the proposed research. First, the un-normalized trained model is used for the evaluation. Next, intermediate layers are chosen to assess the output at that layer. An un-normalized image and a colour normalized image are evaluated at the output layer. This experiment is repeated with a series of test images. Once the output of the layer is produced, the output responses for the un-normalized image and the normalized image are compared using the Jensen-Shannon divergence metric. Overall, this process is repeated for images colour normalized by different methods. Once all methods and images are evaluated, distributions of the Jensen-Shannon divergence metric are compared to examine the colour sensitivity of the model.

C. Colour Normalization

To extend the research of [18], the same colour normalization methods were used in the proposed research. In the context of this paper, colour normalization is used to apply subtle colour changes to the input images to compare the un-normalized model's response between an un-normalized image and a colour normalized image. Due to the numerous normalization methods that exist, the assessment is completed over a variety of colour normalization techniques. The colour normalization methods used in this paper were sourced from the Stain Normalization Toolbox made publicly available by the Department of Computer Science at the University of Warwick [23]. Each method in the toolbox relies on a reference distribution or a image that is applied to the dataset. Non-stain specific methods such as histogram specification ("HS") or colour transfer ("RH") apply the colour statistics of a reference image to a target image in a non-specific manner. This means that colours specific to the H&E stains are not mapped from the reference image to the target image. Rather, the global colour statistics are mapped instead. For stain specific methods, such as non-linear colour mapping ("KH") or spectral matching ("MC"), the colours specific to each stain are mapped accordingly. Stain specific methods are accepted to be more anatomically accurate as nuclei and stromal structures uptake hematoxylin and eosin respectively. Even though the same reference image is used for each colour normalization method, the appearance of the mapped image varies between normalization methods. The last normalization method used in this work was adapted from StainGAN [24] where the implementation was sourced from [25]. The StainGAN is a cycle trained general adversarial network (CycleGAN). Rather than a single reference image, a reference distribution of images is mapped to a target distribution. For this method, it is unknown if the output follows the non- or stain specific ideology. It is expected that the response of the model to each differently normalized image will differ indicating the model's colour sensitivity. The following sections briefly outline each normalization method and their advantages and disadvantages.

1) *Histogram Specification*: Histogram specification or histogram matching is a process where the histogram of a target

image is transformed to that of a reference image [26]. In this process, the transformation to a specified histogram is found. For a red, green, blue (RGB) image the transformation is applied to the individual channels. Histogram specification is known to introduce image artefacts into the resulting image. This is because of that fact the specification is applied to the individual channels. Furthermore, because it is not a stain specific method, tissue structures such as lipids or other structures may be discoloured. While this is true, histogram specification is fast and easy to implement.

2) *Colour Transfer*: Colour transfer was originally applied for natural image colour correction where both the reference image and the target image are converted to the lab space and the *mean* and *standard deviation* of each channel of the reference image is applied to the target image [27]. For natural images, a correction method such as this is useful when the characteristics of an image are desired for images with discolouration. Colour transfer has been adapted in digital pathology using the same concept. The statistics of an image with ideal stain concentration, contrast, and lighting conditions are transferred to that of a target image. In the toolbox, colour transfer is applied using the original method by Reinhard et al, and through a non-linear mapping approach that also uses stain deconvolution [14] [27].

a) *Stain deconvolution*: is a transformation used to pre-process H&E images where an input RGB image is transformed to the stain space. In this space, the resulting image is reflective of the specific stains contained in the tissue specimen. Each channel of the resulting image represents each stain. In images that only have two stains, the third channel represents the error or the colour not captured during the transformation. Stain deconvolution was first introduced by [28] and utilizes the Beer-Lambert law in order to separate the stains:

$$I_p = I_O e^{-\epsilon_p c_p} \quad (1)$$

where I_p is the source image, I_O is the background brightfield, c is the concentration of dye, and ϵ is the molar absorption coefficient. The equation can be remodeled to:

$$OD = -\ln \frac{I_p}{I_O} = \sum_p \epsilon_p c_p = VS \quad (2)$$

$$S = OD * V^{-1} \quad (3)$$

where OD is the optical density values for each channel in the RGB space, V is the stain vector, and S is the concentration of each stain. Estimating the stain vector, V , allows for the estimation of the pure stains contained in the image [29] [28]. However, estimating the stain vector accurately is a challenging task and effects the appearance of the output image. In the normalization toolbox, Khan et al estimates the stain vector and applies the mapping using a non-linear method.

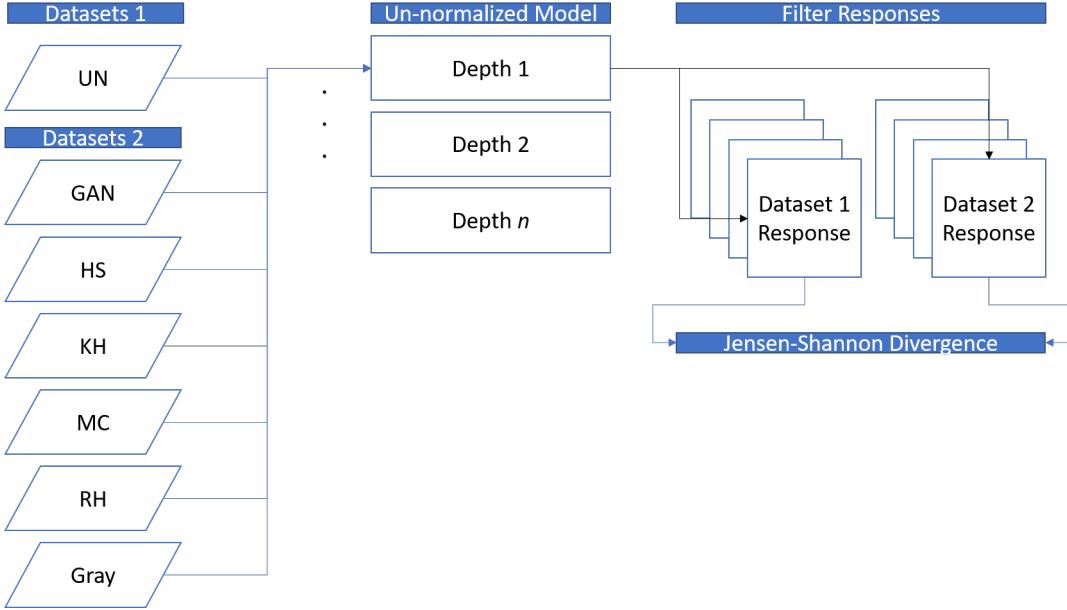


Fig. 1. Experimental design

b) Non-linear mapping: Unlike the traditional colour transfer method, Khan et al proposed a method that is stain specific, uses a supervised method for stain vector estimate, and applies the mapping of a reference image to a target using non-linear mapping. Stain vector estimation is done using two phases of learning and evaluation. During the learning phase stain colour descriptors are derived from training image histograms. This step results in a sparse representation of colour. Next, the sparse representations are used to generate the stain matrix using a probabilistic approach. Once the stain matrix is estimated, the stains from the reference image are mapped to the target using spline based mapping [18] [14]

3) Style-transfer Using Cycle Generative Adversarial Network: Cycle Generative Adversarial Network for style-transfer, or StainGAN, was adapted by [24] for digital pathology applications. Different from a regular GAN for style transfer, StainGAN employs the cycle consistent methodology where instead of a single reference image, an entire reference domain is used to transfer style to a target domain [18] [24]. The cycle consistent methodology describes that if a generator, G , can map domain X to domain Y :

$$G_X : X \longrightarrow Y \quad (4)$$

then generator, F , can map domain Y to domain X :

$$F_Y : Y \longrightarrow X \quad (5)$$

For the cycleGAN framework, the generators are trained to generate images of the opposite domain, while the discriminators confirm if the output images come from the real domain. This cycle methodology is made possible by both a *cycle loss*, L_{cycle} and the adversarial loss, L_{adv} :

$$L = L_{cycle} + \lambda * L_{adv} \quad (6)$$

D. Nuclei Segmentation Using DCNNs

The digital pathology application in which the effect of colour is analyzed is nuclei segmentation. In this paper, the outputs at intermediate layers of the UNET architecture is examined. The original UNET [30] and its variations are popular architectures used in cell segmentation.

The UNET is comprised of a down-sampling path, an up-sampling path, and *skip connections* that connect the two at various depths. These skip connections are known to improve segmentation by including low level features from the down-sampling path in the output at each decoder. A high-level representation of the UNET is depicted in Figure 2. For the purposes of this paper, only baseline UNET architecture is used, and the weighted loss function excluded from the implementation.

E. Jensen-Shannon Divergence

The Jensen-Shannon Divergence (JSD) inequality states that: $\pi_1, \pi_2 > 0$; $\pi_1 + \pi_2 = 1$, where π_1 and π_2 are weights of two probability distributions, p_1 and p_2 . The Jensen-Shannon divergence measure is defined as:

$$JS_\pi(p_1, p_2) = H(\pi_1 p_1 + \pi_2 p_2) - \pi_1 H(p_1) - \pi_2 H(p_2) \quad (7)$$

In this paper, the JSD metric is used to quantify the similarity of the model's response to an un-normalized image and a colour normalized image. At intermediate layers of the model, a prediction, $P(x,y,i)$ can be obtained based off a given input image, $I(x,y,3)$. For un-normalized images, the prediction is represented by $P_{UN}(x,y,i)$ and

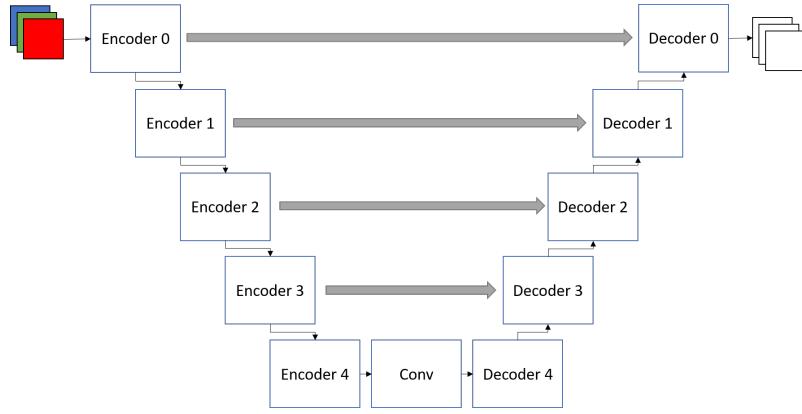


Fig. 2. U-Net Architecture

the prediction of a colour normalized image is $P_{CN}(x, y, i)$ where x , y , and i indicate x-dimension, y-dimension, and filter dimension respectively. Individual filters for $P_{UN}(x, y, i)$ and $P_{CN}(x, y, i)$ where $i \in \{0, 1, 2, \dots, N\}$, and N is the number of filters at a layer, are considered as p_1 and p_2 respectively. Therefore, according to Jensen's inequality, $JS_\pi(P_{UN}(x, y, i), P_{CN}(x, y, i))$ is nonnegative and equal to zero when $p_1 = p_2$. Therefore, the predictions, $P_{UN}(x, y, i)$ and $P_{CN}(x, y, i)$, are more similar when $JS_\pi(P_{UN}(x, y, i), P_{CN}(x, y, i))$ values are closer to 0 and are more dissimilar when $JS_\pi(P_{UN}(x, y, i), P_{CN}(x, y, i))$ values are closer to 1 [31].

For each image used to test the model, a JSD metric is generated. Once all test images are evaluated, box-plots are used to depict the distribution of values across each normalized dataset that were evaluated against the un-normalized dataset. It is expected that normalization methods that contained greater JSD values are indicative of greater colour sensitivity compared to methods which demonstrate lower JSD values. In addition, because we are extending the experiment to different model depths, colour sensitivity can be analyzed as model depth changes. This indicates at which level the model could be most sensitive to colour providing greater insight to the conclusions made by [18].

F. Inverting Deep Representations

To qualitatively investigate how colour is perceived by DCNNs, predictions at various model depths from the previous experiment are inverted. The implementation used in this paper is based on the work of [32]. [32] work on deep image representation inversion was applied to natural images and utilized an implementation of the AlexNet architecture. In addition, [32] hypothesizes that representations collapse irrelevant differences in images; as such, by reconstructing images from various model layers, it may be possible to visualize what kind of features, such as colour, were important to the model at certain depth. By doing so, not only can colour sensitivity be analyzed, but the correlation to what the DCNN perceives can be examined as well.

The inversion of deep representations by [32] was formulated by modeling the problem as a method to solve for the inverse based on an image prior. For instance, given a representation function $\phi_0 : \mathbf{R}^{H \times W \times C} \rightarrow \mathbf{R}^d$ and a representation $\phi_0 = \phi(x_0)$ to be inverted, the reconstruction finds the image $x \in \mathbf{R}^{H \times W \times C}$ that minimizes:

$$x^* = \arg \min(l((\phi(x), \phi_0) + \lambda R(x)) \quad (8)$$

where the loss l compares the image representation $\phi(x)$ to the target ϕ_0 , and R is a *total variation* regulariser that utilizes the image prior. The loss function is simply the euclidean distance between $\phi(x)$ and ϕ_0 :

$$l(\phi(x), \phi_0) = \|\phi(x) - \phi_0\|^2 \quad (9)$$

a) Total Variation Regulariser: The total variation regulariser computes the sum of the absolute differences for neighboring pixel-values in the input image $x + x^*$, where x^* is the estimated reconstruction at a training step. For more details about total variation regulariser please refer to [32].

III. EXPERIMENTAL RESULTS

A. Colour Normalization

a) Colour Normalization Toolbox: In this work, the colour normalization toolbox was used to apply conventional colour normalization methods to the TCGA dataset. Each method of the toolbox required a reference image which is visualized in **Figure 3**. Note, that for the CycleGAN implementation, rather than a single reference image, a dataset comprised of images cropped from the image in **Figure 3** were used for cycle generative adversarial training.

b) CycleGAN: In this work the CycleGAN was adapted from an open source library [25]. It employs a U-Net structured generator and a discriminator network of a combination of convolutional layers, leaky-ReLu activation, as well as instance normalization. λ was set to 10, as in the original implementation, and the learning rate was set to 0.0002 with a batch size of 1. In addition, the CycleGAN implementation utilized an identity loss of $0.1 * \lambda$.

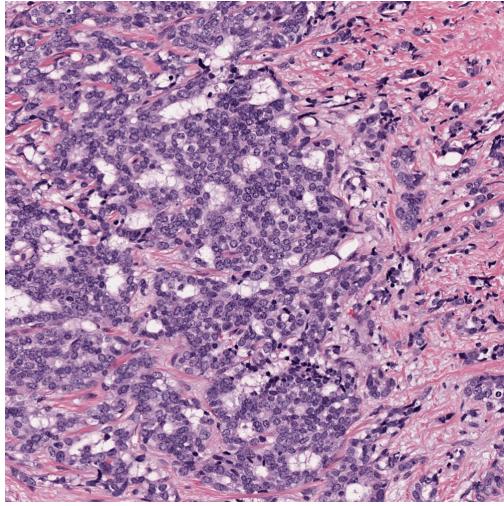


Fig. 3. Reference image for colour normalization toolbox

B. DCNN Architecture Training

1) *Architecture*: Different from the original paper by [18..], only the model based on un-normalized data is trained. This is so the response of the un-normalized model to un-normalized images and colour normalized images can be quantified. A high level description of the U-Net architecture employed in this work can be seen in **Table 1**. Five encoders and decoders are used in the down-sampling and up-sampling paths respectively. The input to the network is an RGB image, and the output is a mask of nuclei, background, and boundary labels which are one-hot encoded.

TABLE I
U-NET ARCHITECTURE

Layer Name	Input Dimension	Output Dimension
Input image	-	-
Encoder network	256 x 256 x 3	8 x 8 x 512
Center	8 x 8 x 512	1 x 1024
Decoder network	1 x 1024	256 x 256 x 32
Output layer	256 x 256 x 32	256 x 256 x 3

2) *Data Preparation and Training Protocol*: The TCGA dataset contains 29 region of interest images (ROIs) of size 1000 x 1000. To increase the amount of data seen by the network, the ROIs are cropped into a total of 464 patches with zero-padding of size 256 x 256. The data split protocol follows approximately 63 % training, 13 % validation, and 24 % testing. During training, data augmentations such as 90 degree rotation, horizontal flipping, and vertical flipping are applied. The model was trained for 40 epochs where the model with the highest validation accuracy was saved. Lastly, the learning rate was set of 0.001 and the optimizer that was used was *Adam*.

C. Colour Normalization Results

Figure 4 depicts examples of images before and after colour normalization was applied. Regarding target image 1, for RH

and HS methods, the colour from the reference image was effectively applied to the target image. However, for MC, KH, and GAN, the resulting colour does not closely resemble the reference image. For example, MC output seems to be quite vibrant. Furthermore, for both KH and GAN, the output images are closer towards the purple spectrum of colour of the reference image. Analyzing target image 2, all methods, with the exception of GAN, apply colour normalization well. For both stain specific and non-stain specific methods, the outputs are quite similar. Lastly, for test image 3, most output seem to resemble the colours contained in the reference image. However, the GAN output exhibits a decrease in contrast which is an example of a disadvantage of some colour normalization techniques.

While it is important to analyze the qualitative results of colour normalization, for the scope of this work, normalization was merely applied to alter the appearance of images. Of interest, is the effect of subtle colour changes on the response of the model at various depths and layers.

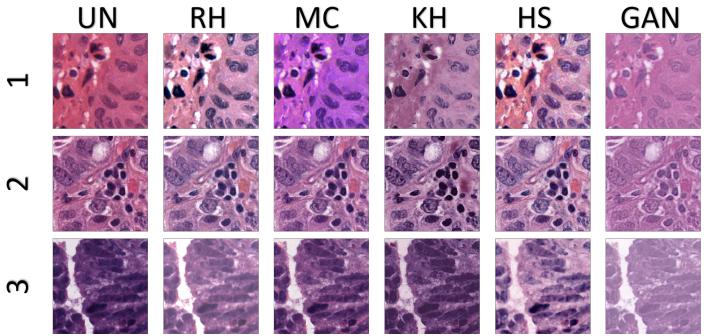


Fig. 4. Examples of colour normalization results

D. U-Net Response to Colour Changes

To understand how subtle colour changes can impact the response of a model tuned to a certain distribution (in this case un-normalized data), 112 sub-patches were used to evaluate the trained model. To reiterate, the response of the model at intermediate layers between an un-normalized image and a normalized image are compared using the Jensen-Shannon divergence metric. This process was repeated for the number of normalization methods used in this experiment as well as at varying depths. The layers where the intermediate outputs were analyzed and some distributions of JSD values can be seen from **Figure 5**.

Analyzing **Figure 5a** and **5b**, the un-normalized model seems to be least sensitive to colour changes evoked by MC normalized images, as indicated by the lowest average JSD. In contrast, GAN normalized images evoke the highest sensitivity as demonstrated by the greatest mean JSD. Note, that the shallowest convolution layer is *conv2d*, whereas, the deepest convolution layer is *conv2d₂₀*. Even at these two depths the sensitivity to the model does not exhibit a large change. This could be an indication that the model has little sensitivity to colour at either depth. Furthermore, the CN methods which

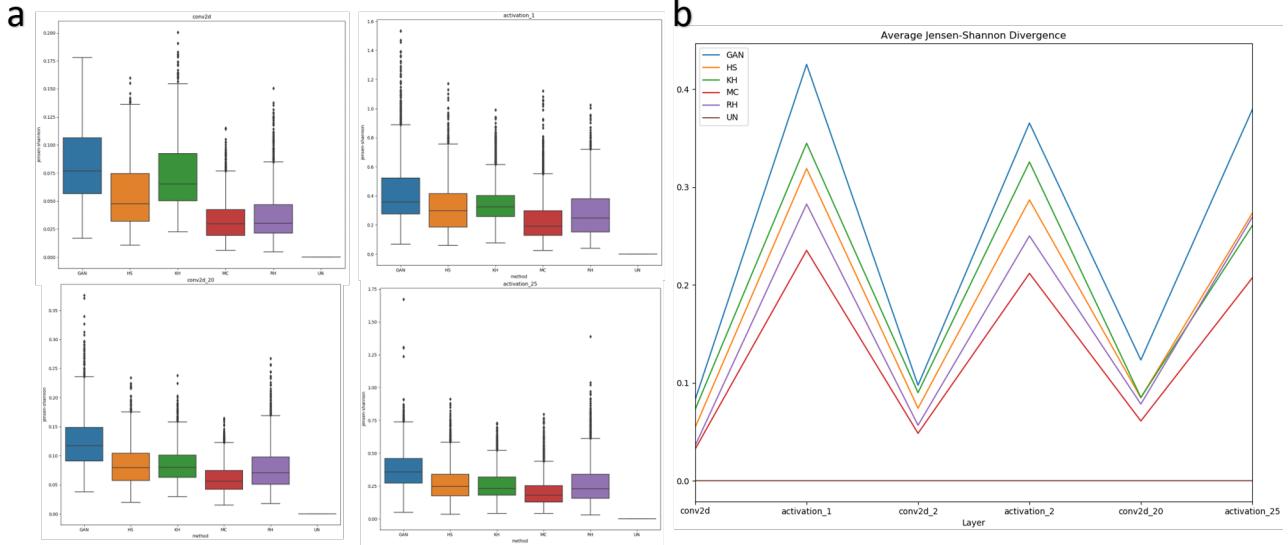


Fig. 5. Jensen-Shannon Divergence At Various Layers

demonstrate a low JSD value could also indicate that the un-normalized and normalized images are very similar in appearance, and therefore evokes a similar response from the model.

E. Inverting Deep Representations

In this paper, the deep representation inversion implementation was adapted from [33]. To further understand how subtle colour changes impact the perception of DCNNs, representations were reconstructed at some layers used in the previous experiment. Analyzing features from the reconstruction could be indicative of what sort the representation is needed to semantically segment nuclei. If the colour of the original image is maintained in the reconstruction, it can be perceived that colour is important for the segmentation task. **Figure 6** depicts the reconstruction of representations at the shallowest layer *conv2d* and a deeper layer *conv2d₂₀*.

Analyzing the reconstructed images, there are not too many differences between those reconstructed from the shallow layer and the deep layer. Furthermore, appearances do not differ greatly between colour normalization methods. This is likely a result of the similar appearance to the un-normalized image even after normalization. The reconstructed image that has noticeable differences from the other methods is the image reconstruction from the KH normalized image. For this image, the stroma background seems to be more vibrant and demonstrates smoother texture.

From the normalized images, the GAN normalized image appears to have the worst appearance, exhibiting low contrast and uniform colour between the nuclei and stroma. It would be expected that nuclei would be more difficult to discern by the model. However, at both intermediate layers, while not quite as distinguishable as the other outputs, the nuclei are still quite discernible. This could be an indication that, in general, the subtle colour changes are not having a large impact on DCNN

segmentation performance, and could explain why in the work of [18] performance did not greatly improve for normalized datasets as expected.

IV. DISCUSSIONS AND CONCLUSIONS

This paper investigated the the role of colour in deep neural network based nuclei segmentation frameworks by quantifying the sensitivity of a trained model to subtle changes in input colour, and by reconstructing deep representations at various depths. In the first experiment, images normalized by different methods and un-normalized images were input to the model simultaneously. From this experiment, it was observed that the MC normalized images elicited a more similar sensitivity to the un-normalized dataset as indicated by the low mean JSD value even at varying depths. This observation could mean that even at shallow and deeper layers colour as a feature is not as important. However, it is interesting that for GAN images the JSD distribution was quite variable at each depth that was evaluated. Perhaps, this could be a result of the lack of contrast in GAN normalized images. It seems as though rather than *color*, contrast might be a greater factor for how the model responds to appearance changes.

For the second experiment, it was observed that there was not a great difference in reconstruction images between depths, as well as, between normalized images. Again, reconstruction images differed slightly for the GAN and KH reconstructions. For GAN, contrast could be an issue as observed from the previous experiment, and for KH perhaps the CN method alters features of the images that the model would otherwise recognize if CN was not applied. For this experiment, a limiting factor could be the ability to invert the image based on the method utilized. The observation that all images are a greenish colour could indicate that the method for inversion is limited to a certain solution, or that the RGB channels are optimally weighted such that the overall output is green. In general, more

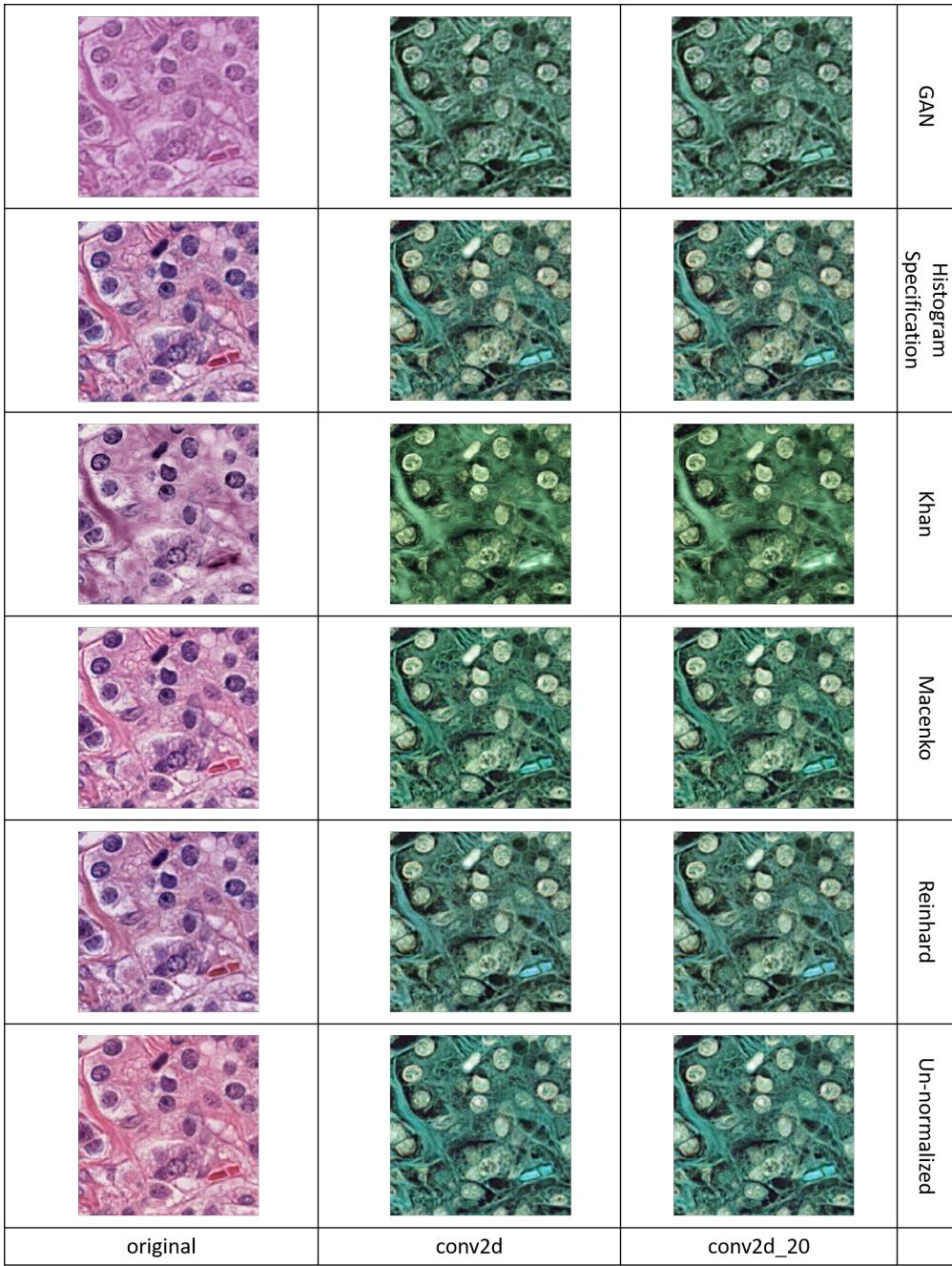


Fig. 6. Reconstruction of intermediate representations at various layers

work and other reconstruction methods should be considered to investigate the perception of colour by DCNNs.

V. ACKNOWLEDGEMENTS

Thank you to Dr. Bruce for providing the opportunity to further my research in this topic.

REFERENCES

- [1] S Walters, C Maringe, J Butler, B Rachet, P Barrett-Lee, J Bergh, J Boyages, P Christiansen, M Lee, F Wärnberg, and et al. Breast cancer survival and stage at diagnosis in australia, canada, denmark, norway, sweden and the uk, 2000-2007: a population-based study. *British Journal of Cancer*, 108(5):1195–1208, 2013.
- [2] G Hortobagyi. *AJCC Cancer Staging Manual*. The American College of Surgeons, 8 edition, 2017.
- [3] Reis-Filho J.S. Baehner F. et al. Rakha, E.A. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res*, 207, 2010.
- [4] April Khademi. Image analysis solutions for automatic scoring and grading of digital pathology image. *Can. J. Pathol*, 5:51–55, 2013.
- [5] F. Cloppet and A. Boucher. Segmentation of overlapping/aggregating nuclei cells in biological images. In *2008 19th International Conference on Pattern Recognition*, pages 1–4, Dec 2008.
- [6] Santa Di Cataldo, Elisa Ficarra, Andrea Acquaviva, and Enrico Macii. Automated segmentation of tissue images for computerized ihc analysis.

- Computer methods and programs in biomedicine*, 100 1:1–15, 2010.
- [7] Nicolas Loménie and Daniel Racoceanu. Point set morphological filtering and semantic spatial configuration modeling: Application to microscopic image and bio-structure analysis. *Pattern Recogn.*, 45(8):2894–2911, August 2012.
 - [8] Mahmudul Hasan and Amit K. Roy-Chowdhury. Continuous learning of human activity models using deep nets. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 705–720, Cham, 2014. Springer International Publishing.
 - [9] B. Parvin, Q. Yang, J. Han, H. Chang, B. Rydberg, and M. H. Barcellos-Hoff. Iterative voting for inference of structural saliency and characterization of subcellular events. *IEEE Transactions on Image Processing*, 16(3):615–623, March 2007.
 - [10] Mitko Veta, Paul J. van Diest, Robert Kornegoor, André Huisman, Max A. Viergever, and Josien P. W. Pluim. Automatic nuclei segmentation in he stained breast cancer histopathology images. *PLOS ONE*, 8(7), 07 2013.
 - [11] Jun Xu, Xiaofei Luo, Guanhao Wang, Hannah Gilmore, and Anant Madabhushi. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*, 191:223, 26/2016 2016.
 - [12] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A. W. M. van der Laak, , and the CAMELYON16 Consortium. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, 318(22):2199–2210, 12 2017.
 - [13] Zahraa Al-Milaji, Ilker Ersoy, Adel Hafiane, Kannappan Palaniappan, and Filiz Bunyak. Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in he images. *Pattern Recognition Letters*, 119:214–221, 2017.
 - [14] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering*, 61(6):1729–1738, June 2014.
 - [15] Jiayun Li, William Speier, King Chung Ho, Karthik V. Sarma, Arkadiusz Gertych, Beatrice S. Knudsen, and Corey W. Arnold. An em-based semi-supervised deep learning approach for semantic segmentation of histopathological images from radical prostatectomies. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 69:125–133, 2018.
 - [16] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, 36(7):1550–1560, July 2017.
 - [17] Maxime W. Lafarge, Josien P. W. Pluim, Koen A. J. Eppenhof, and Mitko Veta. Learning domain-invariant representations of histological images. *Frontiers in Medicine*, 6:162, 2019.
 - [18] Justin Tyler Pontalba, Thomas Gwynne-Timothy, Ephraim David, Kiran Jakate, Dimitrios Androutsos, and April Khademi. Assessing the impact of color normalization in convolutional neural network-based nuclei segmentation frameworks. *Frontiers in Bioengineering and Biotechnology*, 7:300, 2019.
 - [19] Quanshi Zhang and Song-Chun Zhu. Visual interpretability for deep learning: a survey. *CoRR*, abs/1802.00614, 2018.
 - [20] Y. Xu, Z. Jia, Y. Ai, F. Zhang, M. Lai, and E. I. Chang. Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 947–951, April 2015.
 - [21] B. Korbar, A. M. Olofson, A. P. Miraflor, C. M. Nicka, M. A. Suriawinata, L. Torresani, A. A. Suriawinata, and S. Hassانpour. Looking under the hood: Deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 821–827, July 2017.
 - [22] M. Engilberge, E. Collins, and S. Süsstrunk. Color representation in deep neural networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2786–2790, Sep. 2017.
 - [23] Derek Magee. Stain normalisation toolbox, 5 2015.
 - [24] M. Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni. Staingan: Stain style transfer for digital histological images. *CoRR*, abs/1804.01601, 2018.
 - [25] Erik Linder-Noren. Keras: Generative adversarial networks, 2018. Retrieved from: <https://github.com/eriklindernoren/Keras-GANcyclegan>.
 - [26] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing*. Prentice Hall, Upper Saddle River, N.J., 2008.
 - [27] E. Reinhard, M. Adikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, July 2001.
 - [28] Arnout Ruifrok and Dennis Johnston. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*, 23, 01 2001.
 - [29] Marc Macenko, Marc Niethammer, J. S. Marron, David Borland, John T. Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E. Thomas. A method for normalizing histology slides for quantitative analysis. In *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro*, ISBI’09, pages 1107–1110, Piscataway, NJ, USA, 2009. IEEE Press.
 - [30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
 - [31] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, Jan 1991.
 - [32] Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. *CoRR*, abs/1412.0035, 2014.
 - [33] Saurabh. Understanding-deep-image-representations-by-inverting-them. Retrieved from <https://github.com/saketd403/Understanding-Deep-Image-Representations-by-Inverting-Them>.