LUTs GET COLOURING: Optimization of Medical Image Perception in Multiple Modalities Using Colour Mapping and AI

Marie Barberon[1], Hannah Barsouk[2], Naomi Bazlov[3], Anastasia Besier[4], Ari Firester[2], Michael Flynn[5], Alex Miller[3], Ashil Shah[3]

Luxembourg[1], United States[2], United Kingdom[3], Germany[4], Ireland[5],

Mentored by Dr. Vyacheslav Kalchenko
Department of Veterinary Resources
Weizmann Institute of Science, Rehovot, Israel

Abstract

Diagnostic errors in medical image analysis are responsible for 1 in 10 deaths in an outpatient setting and $17-29 billion in wasteful medical spending annually. As the majority of these errors are typically perceptual, developing a mechanism to increase the differentiation of background noise and signal from Regions of Interest (ROI) within medical images would serve to improve patient outcomes. Colour maps were generated using Lookup Tables (LUTs), which were either manipulated within a HSV (Hue, Saturation, Brightness Value) space to resemble parabolic functions, adjusted from pre-existing LUTs from Fiji/ImageJ, or manually segmented to highlight anatomical structures. The efficacy of our generated LUTs was assessed with a survey (n=65) taken by non-radiologists and a Fiji-based automatic lesion detector we coded. While a greater sample size of qualified individuals and generated LUTs would be required to refine our results, five colour maps were shortlisted as being optimal for the detection of brain lesions on T1-weighted MRI scans.

Introduction

A push toward normalizing early screening practices, advancements in understanding and diagnostic tools, and an increase in the accessibility of healthcare and public health resources have all served to revolutionize diagnostic medicine in the 20th and 21st centuries. Furthermore, diagnostic errors in the field of radiology are as high as 40 million per year globally (3-5% of all cases they see). The persistent stagnancy of this number, despite improvements in treatment options and efficacy over the last 100 years, have continued to hinder patient outcomes (Itri et al., 2018).

Radiologists' errors are typically either perceptual (image inspection; this category accounts for around 60-80% of all errors) or cognitive (image interpretation) (Bruno et al., 2015). Procedures which attempt to reduce error include cognitive psychology approaches, such as reducing workload, or educating radiologists on their biases or blindnesses (Brady, 2017). Techniques to rectify perceptual error range from the use of AI to automatically segment legions, to the use of Generative Adversarial Networks (GAN) through neural networks to colourise images (Krupinski, 2010).

HSV colour mapping, in which every colour in an originally low contrast monochromatic image, is mapped by certain values of hue, saturation and brightness value (intensity), can be used to distinguish between the background and regions of interest (ROI) in medical images (Semary, 2018). However, the challenges of colouring medical images include potentially losing anatomical or functional information, over-stimulating radiologists from incorporating too great a range of hues in our colour maps (Kovesi, 2015), and further emphasizing colour perception disparities among those with colour blindness (Geissbuehler, 2013).

This paper used Python to create our analysis code and Fiji/ImageJ to analyse our images. We investigated a multitude of colour mapping techniques, including the generation of parabola-shaped Lookup Tables (LUTs) in HSV space, LUTs which were adjusted from Fiji package contents, and manually generated LUTs which highlighted anatomical structures. We additionally utilized AI to attempt to transfer colour schemes between medical images. The Fiji-based LUTs were determined to be most effective by a survey and lesion-detector, however a larger LUTs and individual sample size would be necessary in the future to corroborate these results.[1]

Materials and Methods

The first aspect of this research was to generate the colour LUTs for application to medical images. We did this through three methods: utilising the conic nature of HSV colour space, using previous findings, and by separating areas of interest in the brain according to their usual gray values.

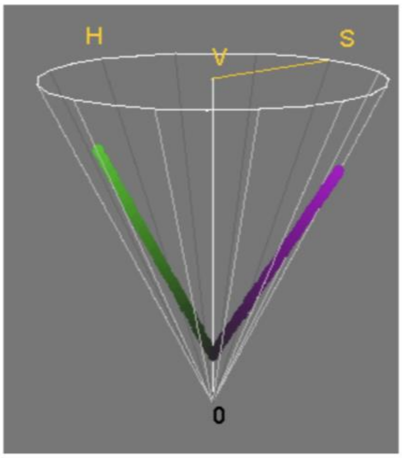


Figure 1: Using HSV colour space to visualize parabolas while generating a LUT

HSV Parabolic LUTs:

Firstly, using HSV space, which is visualised as a cone, LUTs can be produced mathematically by creating a conic section within the cone. The produced section would have different properties depending on the cut produced. When a parabolic shape is produced as shown in Figure 1 (The LUTs displayed here is “ISSI_Cone_6”; refer to Figure 1S). The resulting LUT produced from this parabolic cut has a saturation gradient whereby it repeats from high to low back to high. This aims to remove any outliers from the data of the medical image highlighting the areas of interest. Furthermore, the opposite hue produced from the parabolic arc is effective for colourising medical images.

[1] Due to time constraints, our research focuses on T1-weighted MRI scans of brain cross-sections, although our developed LUTs are inferred to be applicable to other optical imaging modalities.

Adjusting built-in Fiji/ImageJ LUTs:

LUTs were also produced based on previous findings. Matthias Geissbuehler and Theo Lasser found that using magenta green pairings or orange blue pairings worked effectively to show medical information without data loss whilst also functioning for people with red-green colour perception deficiency (Geissbuehler, 2013). Based on these findings we generated several LUTs with similar colour gradients where the ROI would be highlighted.

Manual Structure-Based LUTs:

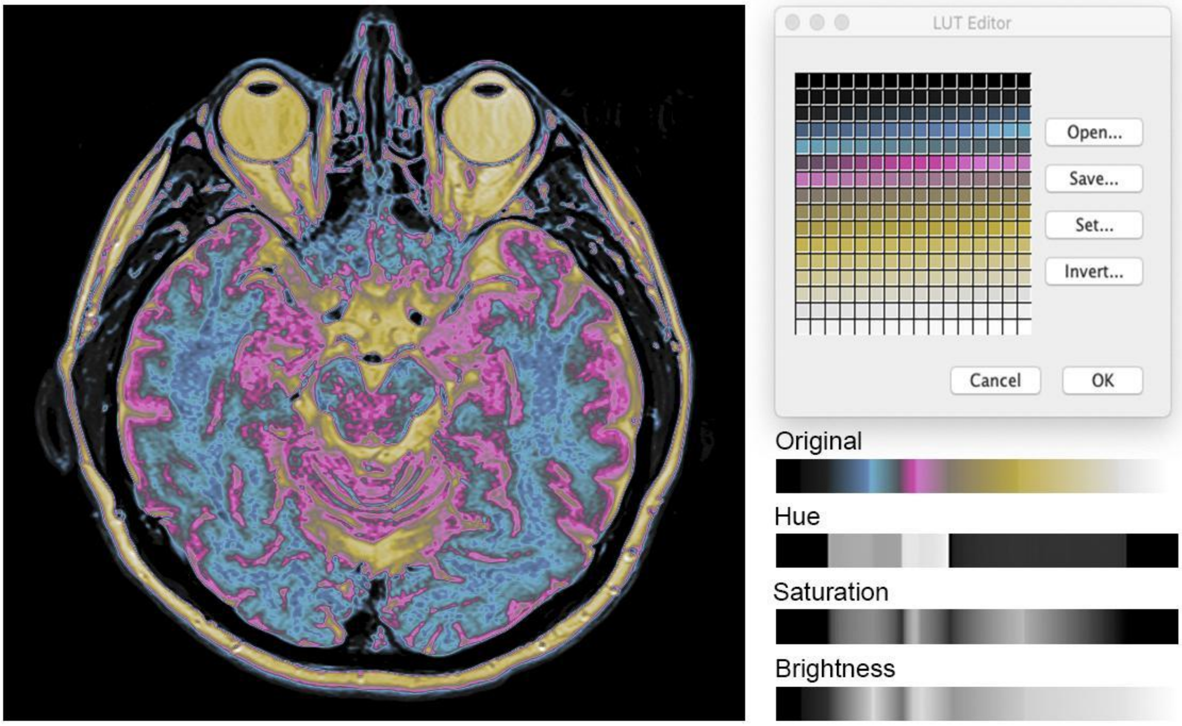
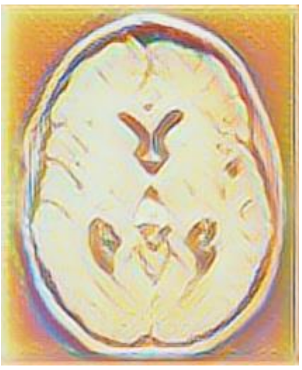


Figure 2: Manually-generated triadic LUTs broken down into hue, saturation, and brightness.

The built-in *Image/colour/Edit LUT* command on ImageJ was applied to a T1-weighted MRI acquired under a creative commons license. Triadic hues (ex: blue, red, and yellow) were applied to entries[2] 48-79, 80-111, and 112-223 in an attempt to denote the gray matter, white matter, and cerebrospinal fluid, respectively, in the image. For each hue, the saturation and brightness values resembled concave down parabolic functions, with the highest values being found at entries 63-64, 95-96, and 167-168. The generated LUT was visualized with the built-in Fiji Plugin *colour Inspector 3D* and 25 new LUTs were generated, each time rotating the hue of the LUT by around 5 degrees. The new LUTs were saved to Fiji's lut folder within *Package Contents* as a .lut files for later use.

Generative Adversarial Networks:

[2] There are a total of 256 entries in a standard Lookup Table (figure 2), each representing a colour with the potential to have a unique hue, saturation, and/or brightness.



We additionally explored the potential of Generative adversarial networks (GANs) to colourise grayscale images. This technique used neural networks to generate new, synthetic data and we used it as a different approach in our project. An online algorithm used our uploaded “style” image and combined it with our “content” image, which was a grayscale brain scan, resulting in a new, colourised version of our brain scan, as shown in Figure 3, and although the results were quite different to our main technique, it provided useful information in our survey expanded upon next.

Figure 3: A GAN colourised MRI brain scan.

Assessing LUTs’ Efficacy:

After generating many unique colour gradients, we used considerations such as minimisation of data loss and being colourblind-friendly to narrow our options down to twenty colour schemes that we thought were more optimal than the others. Subsequently, we sent a survey[3] to ISSI participants to analyse how their perception of scans differed depending on how they were colourised. This was done by writing an HTML page that generated a quiz of 5 questions; one random grayscale MRI, and one random image colourised by each of existing LUTs, a parabolic LUT, a GAN and manual highlighting. Participants of the quiz clicked on the image in the region they suspected an abnormality, and the coordinates of their clicks were logged.

To test each LUT’s effectiveness from a computer’s perspective, we developed a code[4] that automatically searches the picture for artificially made lesions. This code had an edge detector[5] that would convert the colourised brain scan to a grayscale image with accentuated edges and then turn unnecessary noise white (primarily brain folds) that could confuse the subsequent lesion detector. When it found a lesion, the code marked it with a green box; the more lesions found, the more effective the LUT. The two sections are explained in more detail below.

After making the colourised scan black and white, the program needed to find the edges. It combined two filters that determined the vertical and horizontal changes in intensity between pixels. It then replaced each pixel with one that matched that pixel’s relative intensity change as determined by the filters. All pixels that did not fit the threshold were turned white. The lesion detector cycled through every pixel in the picture making one in every 15 an origin. From each origin, it searched in all directions for a nearby border, stopping if it did not find one. To qualify, edges had to be within the

[3] The code for the survey and the images we used can be found at: https://github.com/littlecinnamonroll/littlecinnamonroll.github.io

[4] See code here: https://github.com/AriFirester/AJFCodes/blob/main/Lesion%20Detector

[5] This edge detector code was taken and altered from an already written code from a Princeton CS class website. https://introcs.cs.princeton.edu/java/31datatype/EdgeDetector.java.html

intensity threshold, within the maximum radius, and the change in radius could not exceed a threshold relative to that of the previous edge found. This process was repeated until a complete lesion was found and boxed, or it did not find a complete border and it then moved to the next origin. The LUT could obscure the lesion in two places. The borders could be too unclear for the edge detector to find, and it would not appear. If the lesion edges were not completely found, the program might not have seen it as a closed shape and missed the lesion. The scan with the most lesions found would show that its LUT was the most effective at revealing abnormalities.

Results

Survey:

The results of the survey showed no notable improvement in detecting abnormalities in images colourised by LUTs as compared to the original grayscale scans (see Figure 4, right). Perhaps the most obvious inference that could be made from the available data was that a minority of people who took the quiz recorded abnormalities in the GAN stylised images. This suggests that in order to apply GANs effectively to these images, more than simply using some image's style on a scan would be required. Regrettably, due to the short timeframe of our project, we did not gather a lot of data from the quiz, so the hypothesis that using LUTs enhances abnormality perception could not be proved nor disproved. In future, we would release the survey to a larger population and for a longer time in order to have a chance at collecting more data - we received 65 responses overall. A future direction would include aiming the survey at radiologists, as opposed to a more general audience. Another possible variable to investigate could be the amount of time it takes a radiologist to spot a tumour or lesion in a grayscale scan as opposed to a coloured one. Time is vital when thousands of images and scans are awaiting a diagnosis, and one of our main aims in using LUTs to colour the scans was to reduce stress and difficulty for radiologists—time taken to spot areas of importance on the image is certainly a factor in this.

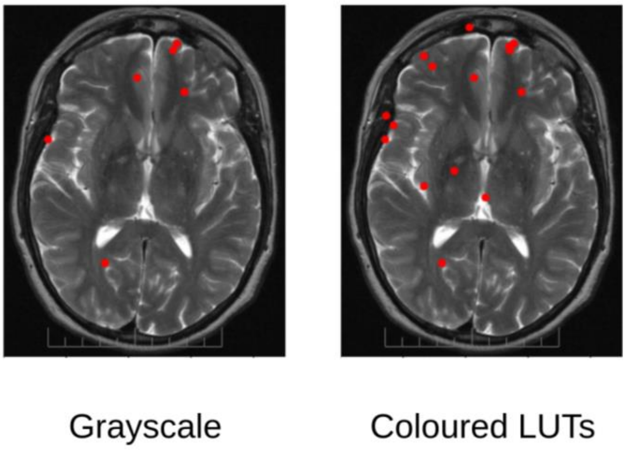
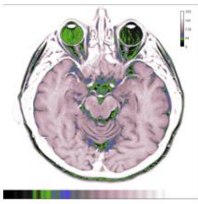
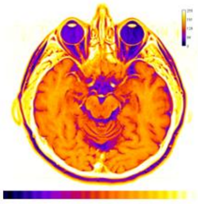
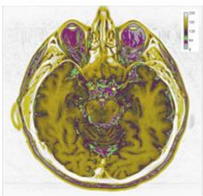
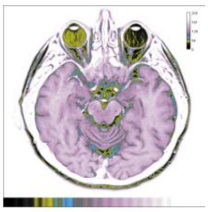
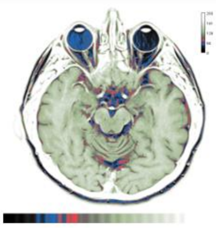


Figure 4: Data points from the survey

Figure 5: LUT scores according to the lesion detector

Name	Picture	Score
Blue green 2		5/5
Fire		4/5
LUT_13b		4/5
Blue yellow 2		4/5
Red blue 2		3/5

Lesion Detector:

After running all the LUTs through the lesion-finding code we developed using a brain scan with 5 lesions, we determined a ranking of the top five LUTS. “Blue green 2” came in first revealing 5/5 lesions with “Fire”, “LUT_13b,” and “blue yellow 2” next with 4/5 each and “red blue 2” with 3/5.

Discussion

In summary, novel colour schemes were generated by using the geometrical properties of the HSV space and by manually editing LUT tables. The effectiveness of the obtained colour maps was then tested by surveying a sample population of individuals in a non-medical context, as well as through a computer program. The results from these analyses were ambiguous, although five colour maps were shortlisted as being optimal for the detection of brain lesions on T1-weighted MRI scans. The best colour schemes proposed were largely in accordance with current literature on optimal colours for visualising grayscale images, but were tested out on a small sample size. Further research could analyse the effects of colour mapping on medical image perception, notably by conducting experiments involving bigger populations, radiologists, or through eye-tracking experiments. As high-quality radiological data becomes increasingly available in the future, machine learning and novel computational techniques should be incorporated with imaging research to continue generating breakthroughs in this field, and improve patient care at the diagnostic level.

Supplementary Information

Scientific Seminar Presentation- Using LUTs for Analyzing Medical Images

Figure 1S (see **Supplementary Data**)

References

Brady, A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable?. *Insights into imaging*, *8*(1), 171–182. https://doi.org/10.1007/s13244-016-0534-1

Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction. *RadioGraphics*, *35*(6), 1668–1676. https://doi.org/10.1148/rg.2015150023

Geissbuehler, M., Lasser, T. (2013). How to display data by colour schemes compatible with red-green colour perception deficiencies. *Optics Express, 21*(8), 9862–9874. https://doi.org/10.1364/OE.21.009862

Itri, J. N., Tappouni, R. R., McEachern, R. O., Pesch, A. J., & Patel, S. H. (2018). Fundamentals of diagnostic error in imaging. *RadioGraphics*, *38*(6), 1845–1865. https://doi.org/10.1148/rg.2018180021

Kovesi, P. (2015). *Good Colour Maps: How to design them.* [Doctoral Dissertation: The University of Western Australia], Cornell University arXiv.org. https://arxiv.org/abs/1509.03700

Krupinski E. A. (2010). Current perspectives in medical image perception. *Attention, perception & psychophysics*, *72*(5), 1205–1217. https://doi.org/10.3758/APP.72.5.1205

Sedgewick, R., & Wayne, K. (2017, October 20). *EdgeDetector.java*. Princeton University. https://introcs.cs.princeton.edu/java/31datatype/EdgeDetector.java.html

Semary, N. A. (2018). A proposed hsv-based pseudo colouring scheme for enhancing medical Image. In D. C. Wyld & D. Nagamalai (Eds.), *Computer Science & Information Technology*: *5th International Conference on Artificial Intelligence and Applications.* (pp. 81–92). AIRCC Publishing. https://doi.org/10.5121/csit.2018.80407

Acknowledgements

A special thank you to Dr. Slava Kalchenko for providing us with the resources and guidance which were necessary to conduct this project, as well as Dr. Aya Shkedy, Dr. Dorit Granot, and Nirit Alon for coordinating the Virtual 2021 ISSI, and allowing the opportunity to participate in this amazing program.

Supplementary Data

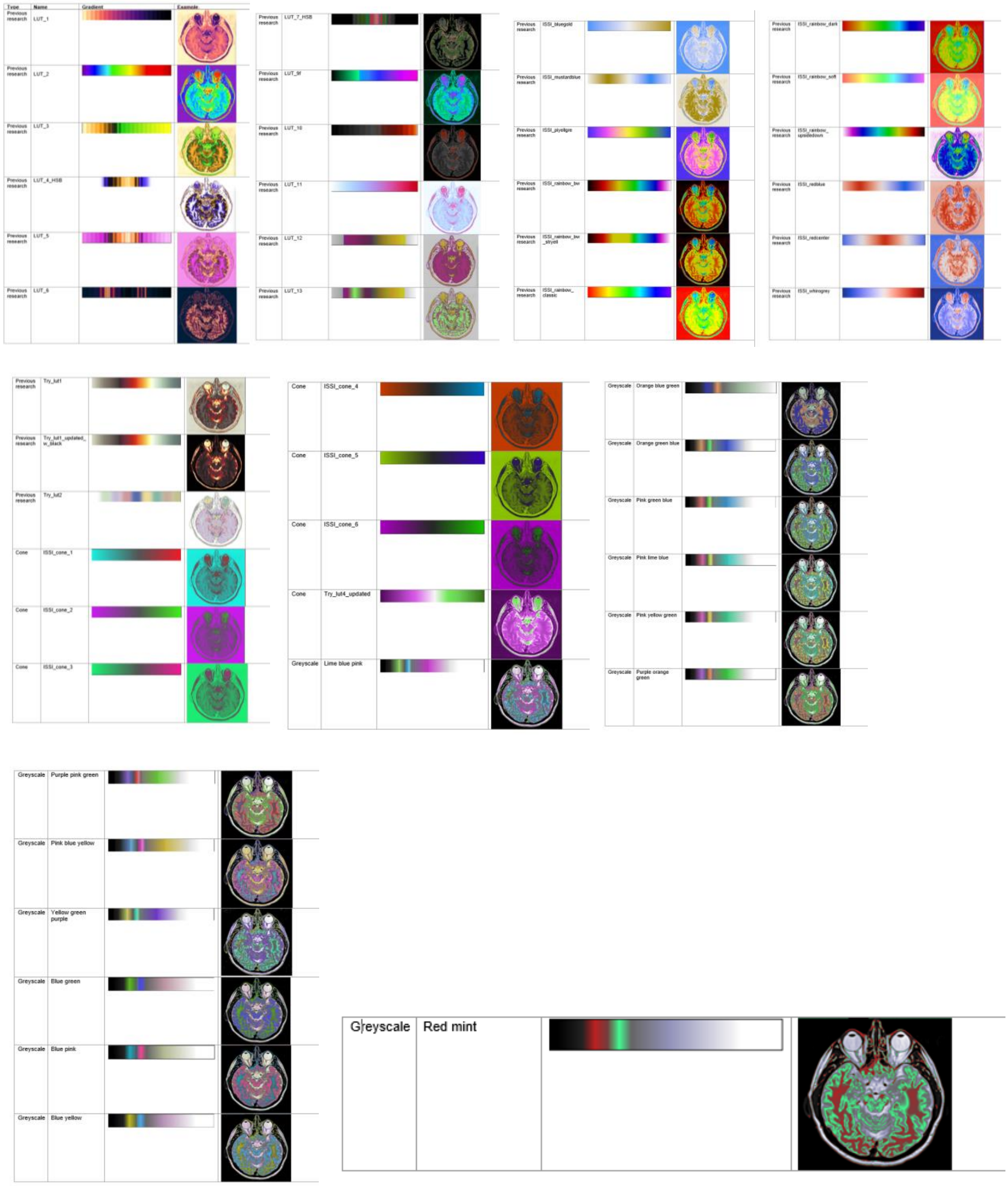


Figure 1S:

54 unique LUTs generated over the course of our project by manipulating the HSV space, adjusting pre-existing LUTs, or manually segmenting grayscale LUTs to highlight anatomical structures.