



Quantitative Performance Evaluation of Commonly Used Colormaps for Image Display in Myocardial Perfusion Imaging: Analysis based on Perceptual Metrics

Miyokardiyal Perfüzyon Görüntülemeye Görüntü Ekranı için Yaygın Olarak Kullanılan Renk Haritalarının Kantitatif Performans Değerlendirmesi: Algısal Metriklere Dayalı Analiz

© Mohsen Qutbi

Shahid Beheshti University of Medical Sciences, School of Medicine, Department of Nuclear Medicine, Tehran, Iran

Abstract

Objectives: To quantitatively evaluate the performance of the most used colormaps in image display using perceptual metrics and to what extent these measures are congruent with the true intensity or uptake of pixels at different levels of defect severity in simulated cardiac images.

Methods: Six colormaps, labeled "Gray", "Thermal", "Cool", "CEqual", "Siemens" and "S Pet" extracted from FIJI ImageJ software are included. Colormap data are converted from the red, green, blue color space to CIELAB. Perceptual metrics for measuring "color difference" were calculated, including difference (ΔE^{76}) and "speed". The pairwise color difference in every two levels or entries is visualized in a 2-dimensional "heatmap distance matrix" for each colormap. Curves are plotted for each colormap and compared. In addition, to apply this technique to clinical images, simulated short-axis cardiac slices with incremental defect severity (10% grading) were employed. The circumferential profile curves of true pixel intensity, lightness or luminance, and color difference are plotted simultaneously for each defect severity to visualize the concordance of the three curves in various colormaps.

Results: In 0% defect, all the curves are at the highest level, except for "s pet", in that the lightness is not at its maximum value. In the phantom with 10% defect (or 90% of maximum value), discrepancies among curves appear. In "Siemens", the ΔE^{76} drops sharply. In "Siemens" colormap, the ΔE^{76} drops sharply. In 80% defect, ΔE^{76} curve, in "gray" colormap drops more slowly than other curves of other colormaps. In "s pet", lightness curve rises paradoxically, although the count intensity and ΔE^{76} curve match. In 70% defect, again, the curves are in good agreement in "thermal", "Siemens" and "cequal". However, a consistent lag exists in "gray". Up to 50% defect, curves maintain their expected pattern, but in defects more severe than 40%, lightness and ΔE^{76} curves in "cool" and "cequal" rise paradoxically, and in "thermal", they start to slow down in descent. In "Siemens", falling pattern of the three curves continues. For "s pet" colormap, an erratic pattern of lightness and ΔE^{76} curves exists.

Conclusion: Of 6 colormaps investigated for estimating defect severity, "grayscale" is less favorable than others and "thermal" performs slightly better. "S pet" or rainbow, which is used traditionally by many practitioners, is strongly discouraged. The "Siemens" colormap suffers from decreased discriminating power in the range of mild to moderate/severe. In contrast, the "cool" and "cequal" colormaps outperform the other colormaps employed in this study to some extent, although they have some shortcomings.

Keywords: Colormap, look-up table, performance, perceptual metric, image display, quantitative analysis

Address for Correspondence: Mohsen Qutbi Asst. Prof., Shahid Beheshti University of Medical Sciences, Faculty of Medicine, Department of Nuclear Medicine, Tehran, Iran

Phone: +0982123031250 **E-mail:** mohsen.qutbi@gmail.com, mohsen.qutbi@sbmu.ac.ir ORCID ID: orcid.org/0000-0002-8347-605X

Received: 23.09.2023 **Accepted:** 21.01.2024 **Epub:** 09.02.2024



Copyright© 2024 The Author. Published by Galenos Publishing House on behalf of the Turkish Society of Nuclear Medicine. This is an open access article under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

Öz

Amaç: Görüntü ekranında en çok kullanılan renk haritalarının performansını algısal ölçümler kullanarak niceliksel olarak değerlendirmek ve bu ölçümlerin, simüle edilmiş kardiyak görüntülerdeki farklı defekt şiddeti seviyelerindeki piksellerin gerçek yoğunluğu veya alımıyla ne ölçüde uyumlu olduğunu değerlendirmektir.

Yöntem: Çalışmaya FIJI ImageJ yazılımından çıkarılan “Gri”, “Termal”, “Soğuk”, “CEqual”, “Siemens” ve “S Pet” etiketli altı renk haritası dahil edilmiştir. Renk haritası verileri RGB renk uzayından CIELAB’a dönüştürülür. “Renk farkını” ölçmek için “fark” (ΔE^{76}) ve “hız” dahil algısal ölçümler hesaplandı. Her iki seviyedeki veya girişlerdeki ikili renk farkı, her renk haritası için 2 boyutlu bir “ısı haritası uzaklık matrisinde” görselleştirilir. Her renk haritası için eğriler çizilir ve karşılaştırılır. Ek olarak, bu tekniği klinik görüntülere uygulamak için, artan defekt şiddetine (%10 derecelendirme) sahip simüle edilmiş kısa eksenli kalp kesitleri kullanıldı. Gerçek piksel yoğunluğunun, açıklığın veya parlaklığın ve renk farkının çevresel profil eğrileri, çeşitli renk haritalarında üç eğrinin uyumunu görselleştirmek amacıyla her defekt şiddeti için eşzamanlı olarak çizildi.

Bulgular: %0 defektte “s pet” dışında tüm eğriler en yüksek seviyede olup açıklık maksimum değerinde değildir. %10 defektli (veya maksimum değer %90’ı) fantomda eğriler arasında tutarsızlıklar görünür. “Siemens”te ΔE^{76} keskin bir şekilde düşüyor. “Siemens” renk haritasında ΔE^{76} keskin bir şekilde düşüyor. %80 defektte, ΔE^{76} eğrisi, “gri” renk haritasında diğer renk haritalarının diğer eğrilerine göre daha yavaş düşer. “S pet”te, sayım yoğunluğu ve ΔE^{76} eğrisi eşleşse de açıklık eğrisi paradoksal olarak artar. %70 defektte ise yine “termal”, “Siemens” ve “cequal” eğriler iyi bir uyum içindedir. Ancak “gri” renkte tutarlı bir gecikme mevcuttur. %50’ye kadar defektte, eğriler beklenen şeklini korur, ancak %40’tan daha şiddetli defektlerde, açıklık ve ΔE^{76} eğrileri “soğuk” ve “cequal”de paradoksal olarak yükselir ve “termal”de alçalışırken yavaşlamaya başlar. “Siemens”te üç eğrinin düşme paterni devam eder. “S pet” renk haritası için düzensiz bir açıklık modeli ve ΔE^{76} eğrileri mevcuttur.

Sonuç: Defekt şiddetini tahmin etmek için incelenen 6 renk haritasından “gri tonlamalı” diğerlerine göre daha az avantajlıdır ve “termal” biraz daha iyi performans göstermektedir. Birçok uygulayıcı tarafından geleneksel olarak kullanılan “s pet” veya gökkuşağı kesinlikle önerilmez. “Siemens” renk haritası, hafif ile orta/şiddetli aralığında ayırt edici gücün azalmasından muzdariptir. Buna karşılık “cool” ve “cequal” renk haritaları, bazı eksiklikleri olsa da, bu çalışmada kullanılan diğer renk haritalarından bir dereceye kadar daha iyi performans göstermektedir.

Anahtar kelimeler: Renk haritası, arama tablosu, performans, algısal metrik, görüntü ekranı, niceliksel analiz

Introduction

Image display or visualization is one of the key steps in medical image interpretation after necessary image processing and analysis. In most circumstances, readers apply various combinations of colors to gray-intensity images for various reasons, known as image pseudo-coloring. Visualization is generally enhanced and optimized depending on the purpose and application. Lesion-finding tasks and estimating the level of intensity (for example severity of defects in myocardial perfusion imaging) require the use of different colormaps. The former requires enhancing contrast and maximizing the conspicuity of the area of interest (or lesions), and the latter is based on the input-output relationship between the original data and the displayed image (1,2,3,4). The input-output relationship is related to the intensity transformation of images and is defined as linear or other simple nonlinear mathematical functions (logarithmic or exponential) relating input to output. Except for the direct input-output relationship in the gray colormap, this relationship is non-linear for other colormaps and therefore compromises the estimation of relative intensity. This non-linearity between input and output frequently occurs with multiple-hue colormaps (5,6,7,8). Furthermore, display in two- or three-dimensional modes using different methods of shading and rendering and fusing of images also affects one’s judgment (1,9,10). However, based on the application, particular colormaps may be preferred. Furthermore, colormaps that consist of different hues, saturation, and intensity may have a strong

effect on the reader’s or interpreter’s perception. In most cases, medical imaging practitioners use commercially available colormaps embedded in the software by a vendor and it is used because one is accustomed to using it in their interpretation or based on their own prior experience, preference, or convention. Despite this fact, domain-specific colormaps are widely used across various disciplines (3).

Frequently, colormaps are compared subjectively. Therefore, interpretation is subject to biased estimation because of the different interobserver perceptual impacts of colors. Because of the complex multi-faceted nature of the phenomenon of color perception, quantification and measurement are inherently difficult tasks, and the results may not satisfactorily reflect the true psychological consequences (9,11,12,13,14,15,16,17). Owing to the reasons mentioned and difficulties in quantifying the characteristics, specifications, and behavior of different colormaps, few studies have investigated the properties of various colormaps and, more importantly, their implications in clinical settings such as nuclear medicine imaging. Evaluation of colormaps with more quantitative measures facilitates comparison and enables one to choose the right one. In addition, except for grayscale, colors in other colormaps are perceived differently in humans. This issue complicates the problem because using perceptual metrics and assessing colors in various available color spaces are not straightforward. Employing mathematical modeling for the psychophysical and physiological aspects

of different colors and their association in a sequential pattern may be beneficial to some extent. Of the various color spaces available, CIELAB ($L^*a^*b^*$) and CIELUV are more compliant with human perception of colors with slight differences. In the CIELAB color space, the dominant feature is the perceptually uniform distances between colors. Therefore, the colors in the red, green, blue (RGB) color space can be converted to corresponding ones in the $L^*a^*b^*$ color space to examine the distance of colors in a sequential colormap (2,3,4,18,19). Thus, the evaluation of the compatibility of its pattern with the raw data values of pixels (or counts) in images is feasible. Then, according to the application, one can find the agreement and correlation between them.

The present study aims to quantitatively evaluate the performance of mostly used colormaps in image display and visualization by measures that are considered perceptual metrics. In addition, to examine the extent to which these measures are congruent with the true intensity or uptake of pixels over various values, i.e., different levels of defect severity in simulated cardiac single photon emission computed tomography (SPECT) or positron emission tomography (PET) images.

Materials and Methods

Mathematical Representation and Quantitative Metrics

Mathematically, a colormap is a discrete-valued function that relates the values as input to the corresponding values as output. Discretized input values, because of spatial sampling and quantization, are sorted and then “mapped” to another set of discrete values. For an 8-bit or 256-level colormap function, this process can be expressed mathematically as follows:

$$\begin{cases} F: p_i \rightarrow c_i \\ i \in [1,256]; \end{cases}$$

In this case, index i represents the i th bin of the image histogram and the one-to-one function, F , maps p_i to c_i as elements of domains and codomains. By this means, a distinct color or shade of gray is assigned. For colormaps consisting of a range of color hues, each level or entry is formed by a triple component as channels (R, G, and B representing red, green, and blue). Therefore, c_i equals $[c_i(R), c_i(G), c_i(B)]$. In gray shades, in contrast, all the values in each triplet are equal (4). To demonstrate the overall characteristics and major features of each colormap, a gray transformation is applied using the following formula (20):

$$g(x, y) = \alpha f(x, y, R) + \beta f(x, y, G) + \gamma f(x, y, B)$$

Where the coefficients, $\alpha = 0.2989$, $\beta = 0.5870$, and $\gamma = 0.1140$ are weights according to the perceptual impacts of different wavelengths in humans. f is the original image and g is the gray-transformed image. The resulting gray intensity is graphed and compared.

Next, for each colormap, a conversion is implemented from the RGB color space to the CIE $L^*a^*b^*$ or CIELAB color space. First, the colormap in RGB space is converted to the corresponding one in XYZ color space, and the XYZ/ $L^*a^*b^*$ transformation is conducted. To accomplish this transformation, a MATLAB built-in algorithm is employed. L^* denotes the luminance or lightness and ranges from 0 to 100. Values of 0 and 100 specify black and white in the image. a^* and b^* indicate the perpendicular coordinate axes in the chromaticity plane. The a^* axis determines the amount of red and green hues (in positive and negative directions, respectively). Likewise, the b^* axis determines the number of yellow and blue hues (in positive and negative directions, respectively). For graphing, L^* is demonstrated as a straight or curved line ranging from 0 to 100 as the target line, and the order of the color components in the colormap is depicted as a path (multiple segmented curved lines) in the a^*b^* plane.

To compare the impact of colors in neighboring or distant levels in the colormap the pairwise “color difference” is computed by a distance metric in the Euclidean space as follows:

$$\Delta E_{i,j}^{76} = \sqrt{((L_i^* - L_j^*)^2 + (a_i^* - a_j^*)^2 + (b_i^* - b_j^*)^2)}$$

$$i, j \in [1,256], i \neq j;$$

In that, ΔE^{76} represents the distance between two colors in CIELAB color space (version 1976). This metric is computed between each possible pair in the colormap. For an 8-bit colormap (256 levels or entries and a range of values from 0 to 255 in each level), a heatmap distance matrix with a size of 256×256 is computed, and the value of ΔE^{76} is visualized as gray intensity at the intersection of the row and column of interest. The higher the value of that element in the matrix (pixel), the brighter the intensity of that pixel. Furthermore, the “speed” of color change between two arbitrary levels in the colormap (levels i and j for example) is calculated using equations (4,21):

$$V_{i,j} = \frac{\Delta E_{i,j}^{76}}{l_i - l_j}$$

V_{ij} is the speed between two arbitrary levels (I_i and I_j). Therefore, for "local speed" between two successive levels (I_j and I_{j+1}), the equation is as follows:

$$V_{j,j-1} = \Delta E_{j,j-1}^{76}$$

Then, the average, standard deviation, maximum speed, and minimum speed are calculated (4).

Colormaps

In the present study, 15 colormaps were categorized into four groups according to general similarity in hues included (i.e., the sequential pattern of hues and the gray intensity curves), as shown in Figure 1. Six colormaps that are frequently used for applications of medical visualization were enrolled and analyzed. To perform this task, the original files for generating colormaps were extracted from FIJI ImageJ software, which is a Java-based framework for biomedical image processing and analysis developed by the National Institutes of Health (22). The 6 selected colormaps under investigation in this study are "Gray", "Thermal", "Cool", "CEqual", "Siemens" and "S Pet". As can be seen in the figure, in the upper panel, the colormaps are presented from the lower bound at the leftmost and the upper bound at the rightmost side of each color bar. The gray intensity values of each colormap are plotted as shown in the lower panel. In category 1, the gray intensity curves of "gray" and "thermal" are fitted to the target line nearly perfectly. There is a consistent linear pattern over the entire spectrum. In category 2, "cool", "cequal", "ge" and "mmc" are present, and the patterns of the gray intensity curves are roughly similar. There is a linear pattern from

the mid-range to the upper bound with a steeper than the target line. In the lower half, no consistent relationship exists "cequal" reaches a plateau near the upper bound. In category 3, "Siemens", "hot iron", "warm metal and "fire" are selected. There is an approximately linear pattern from the lower to upper bounds with some irregular fluctuations. Finally, in category 4, "s pet", "rain bow", "a squared", "physics" and "prism" are included. There are wide fluctuations over the entire range of colormaps, which are far from the target line.

These labels are the standard names that are well known by researchers and practitioners in biomedical disciplines, although some differences exist, particularly in non-medical contexts. The files in the format of .lut files are converted to .csv text files and are then imported into MATLAB software for processing and analysis. The file consisted of the values of RGB triples for 256 levels or entries. In each entry, values are in the range [0,255] as integers that indicate the 8-bit format of colormaps. For visualization and analysis, necessary reformatting, transformation, and conversion from RGB to other color spaces are implemented. The scale is preserved as linear in the original data (no logarithmic, exponential, or other custom-function-based transformation is applied).

Cardiac Phantom

A simplistic model of the short axis of the left ventricle (LV) is designed in a 128×128 matrix. Gaussian blurring is then imposed (with $\sigma = 8$ pixels). To simulate a perfusion defect, the short-axis phantom is divided into four sectors, including anterior, septal, lateral, and inferior, each 90°. The defect is placed in the anterior sector with a graded

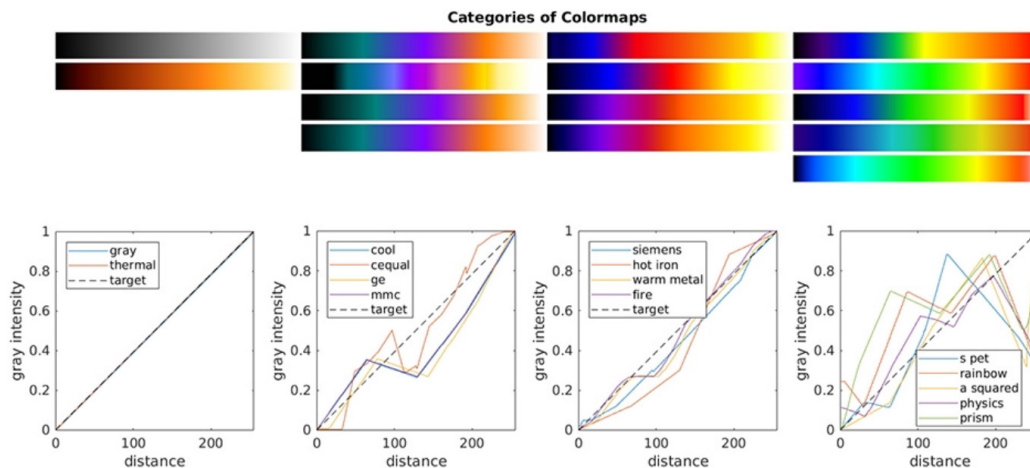


Figure 1. Colormaps are visually presented (upper panel) and the plots of gray intensity (lower panel). In the upper panel, the colormaps are presented from lower bound at the leftmost and the upper bound at the rightmost side of each color bar. The gray intensity values of each colormap are plotted as demonstrated

10-percent increment from no defect (100% of maximal myocardial uptake) to absent uptake (0% of maximal myocardial uptake). In total, 9 phantoms with various levels of defect severity were obtained. This simplistic model is employed because, in cardiac images, the field is limited to the organ of interest (LV).

Tomographic Reconstruction

All phantoms are entered one by one during the process of tomographic reconstruction. Iterative maximum likelihood expectation maximization (MLEM) is used. An acquisition arc of 360° and angular sampling of 3° were set. Collectively, 120 projections were generated during the Radon transform. Then, the generated sinograms are back-projected (inverse Radon transform) to create the tomographic slices. This procedure is repeated 10 times (number of iterations: 10), and in each step, the tomographic slice is updated by dot multiplication of the tomographic slice of the previous iteration to the error image resulting from the element-by-element division of the measured and estimated sinograms. The initial (guess) image is considered to be a uniform one-valued image of the same size as the phantom.

Image Analysis and Plotting

For the analysis of tomographic images, maximum normalization is applied. Thus, the maximum value in each tomographic slice is set to one. This process is performed to ensure the comparability of images and curves and to

avoid possible normalization errors. To visually assess the tomographic images and defects, they are displayed side-by-side using 6 colormaps introduced previously. Each slice with a known perfusion defect in the anterior sector (or wall) is profiled by a vertical line crossing the anterior and inferior walls. The intensity profile is then plotted. To plot the circumferential intensity profile of the myocardium, 16 samples (non-uniformly dispersed spatially in the left and right half of the LV) were selected. Hence, the radial profile of the walls is fitted to a Gaussian curve, and the sample points are chosen such that they reside in the maximum or peak of the radial profile. Then, the real count or intensity profile curve is drawn. In addition, for each colormap and each defect severity, the lightness of the luminance curve and the ΔE^{76} curve (as discussed before) are plotted for those sample points.

Implementation and Coding

Codes for designing phantoms, image acquisition, iterative tomographic reconstruction, and image analysis are written in the MATLAB programming language and run in MATLAB software (The MathWorks Inc., version 2021b). The colormap files (or lookup tables as .lut text file formats) are extracted from FIJI ImageJ software.

Results

In Figure 2, the heatmap Euclidean distance matrix is visualized for 6 colormaps selected for more in-depth analysis. The entire range of the colormap is displayed at

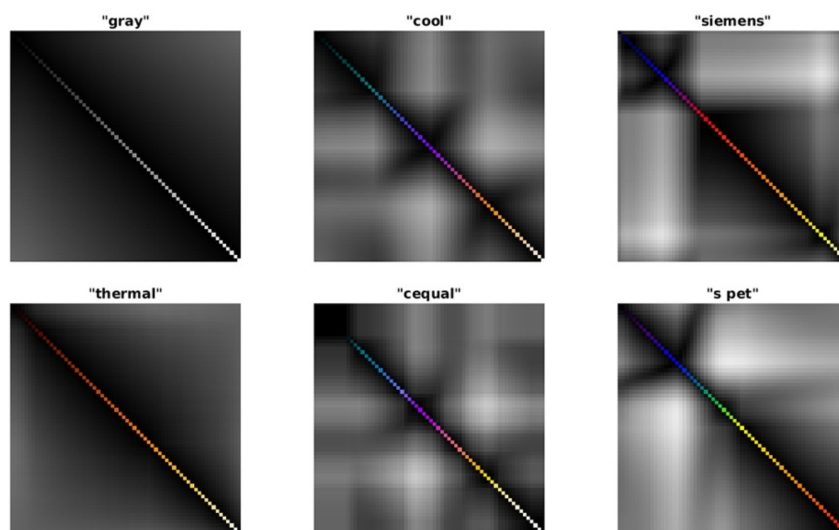


Figure 2. Heatmap Euclidean distance matrices for 6 colormaps are presented. Each panel corresponds to a colormap (here, 6 colormaps). The colormap is displayed in the main diagonal of the matrix, from lower to upper bounds of colormap starting from the left top corner of the matrix to the right bottom corner respectively. The distance between each two pairs of levels or entries (i.e., color hue or shades of gray) is displayed as a shade of gray. The more the distance between two arbitrary pairs on the main diagonal of the matrix, the higher the intensity of the corresponding pixel (which is located at the intersection of the row and column of those pairs)

the main diagonal of the matrix, and the distance between each pair of levels or entries (i.e., color hue or shades of gray) is displayed as a shade of gray. The greater the distance between two arbitrary pairs on the main diagonal of the matrix, the higher the intensity of the corresponding pixel (which is located at the intersection of the row and column of those pairs). As can be seen in the "gray" colormap, the maximum distance is between the lower and upper bounds, as expected. The distance between neighboring pairs (that are close to each other) is too small to be recognized (barely distinguishable) by the human eye. In addition, there is a consistent pattern of distance over the entire range. In "thermal" colormap, an almost similar pattern exists with the exception that the distance between the lower and upper bounds is more prominent than "gray" colormap. In other words, the distance between each pair is augmented in "thermal" compared to "gray". In other colormaps, "cool", "Siemens", "cequal" and "s pet", no regular pattern is visible. In "cool" colormap, the highest distance between pairs is in the mid-range of the map. The largest distance is between the orange and blue pixels. In the upper third of the spectrum, the distance is minimal. An almost similar pattern is evident in "cequal". In "Siemens", the lower third (blue pixels) is remarkably distant from the upper two-thirds (red-orange-yellow pixels), but the pixels in the upper two-thirds of the spectrum are minimally distant. In "s pet" the greatest distance is between the lower and middle thirds (blue pixels and green-yellow pixels).

The curves of lightness or L^* in the CIELAB color space and ΔE^{76} for 6 colormaps are provided in Figure 3 and Figure 4, respectively. As demonstrated in Figure 3, the y-axis and x-axis denote the absolute value of lightness in the CIELAB color space and the levels or entries from 1 to 256. The

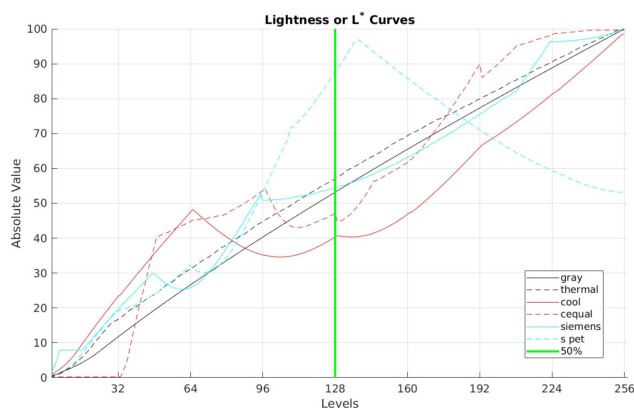


Figure 3. Lightness or L^* curves. The y-axis and x-axis denote the absolute value of lightness in CIELAB color space and the levels or entries from 1 to 256. The mid-value (50% of myocardial uptake) is indicated by a vertical green line in bold and splits the graph into two parts (lower half and upper half)

mid-value (50% of myocardial uptake) is indicated by a vertical green line in bold and splits the graph into two parts (lower half and upper half). The curves of "gray" and "thermal" are straight lines almost fitted to the target line. The curve of "Siemens" follows the target line to a high extent. The curve of "s pet" is far deviated from the target line. In the upper half (myocardial uptake in the defect 50% to 100% of maximal myocardial uptake), the curves of "cool" and "cequal" rise almost linearly, but in the lower half, an irregular pattern is evident. In the upper half of the graph (>50%), the steepest line is for "cool" but for "gray", "thermal" and "Siemens", the slope is lower and equals that of the target line. The curve for "s pet" is reversed. In Figure 4, the y-axis and x-axis denote the absolute value of the color difference in the CIELAB color space and the levels or entries from 1 to 256. The mid-value (50% of myocardial uptake) is indicated by a vertical green line in bold and splits the graph into two parts (lower half and upper half). Here, the ΔE^{76} is the difference between the color in the uppermost bound and the color of an arbitrary level so that the value in the rightmost part of the x-axis ends in 0. For "gray" spectrum, the ΔE^{76} curve is a straight line starting from max at level 1 to 0 at level 256. In "thermal" colormap, the curve plateaus in the first half (<50%) and gradually falls in the upper half (>50%). The curves of "cool" and "cequal" colormaps are similar in pattern, but the output range is wider compared to "thermal" (0-100 for "thermal" and 0-140 for "cool" and "cequal"). In the upper half, the curves are almost a straight line with a notch in the mid-way (about 75% of maximal uptake or 25% defect severity). The ΔE^{76} curve for "Siemens" shows an irregular pattern in the lower half and remains steady in the upper half before a sharp decline at the end of the

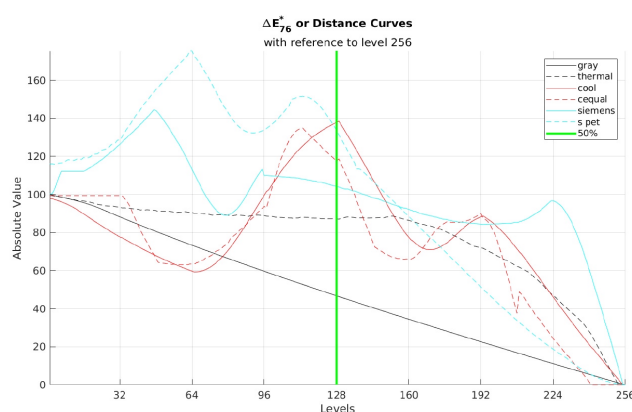


Figure 4. ΔE^{76} curves for the 6 colormaps. The y-axis and x-axis denote the absolute value of color difference in CIELAB color space and the levels or entries from 1 to 256. The mid-value (50% of myocardial uptake) is indicated by a vertical green line in bold and splits the graph into two parts (lower half and upper half)

range. The curve is plateau at a wide range in the middle. A similar scenario is for "s pet" in the lower half but falls in a linear pattern in the upper half.

The local speed and its statistical measures for each colormap are shown in Figures 5, 6. In Figure 5, the blue line indicates the original data, and the red curve in bold indicates the smoothed data. For "gray", the speed is high in the leftmost part of the range but uniformly decreases then. In absolute terms, the changes in speed are profoundly low compared with those of other colormaps. In "thermal", the speed is remarkably high in the rightmost and leftmost parts of the range. For "cool" and "cequal"

colormaps, the speed is minimum in the upper and lower zones of the range, but in the middle, it demonstrates several peaks. For "Siemens", the speed curve is irregular over its entire range. Finally, the speed is uniform in the upper half and peaks sharply.

The tomographic slices reconstructed using the iterative MLEM method with graded defect severity are displayed in Figure 7. Each row is dedicated to one colormap. The incremental defect severity of the anterior wall was calculated as the percentage of maximal myocardial uptake (100% means no defect and 0% means no uptake). Here, the defects based on the color that appears to the

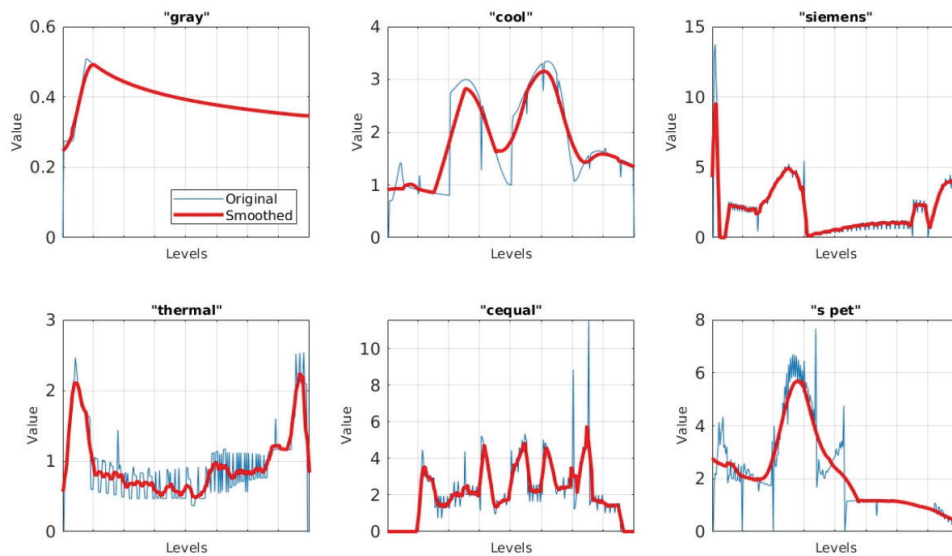


Figure 5. The local speed is displayed for each colormap. The blue line indicates the original data and the red curve in bold in the smoothed data

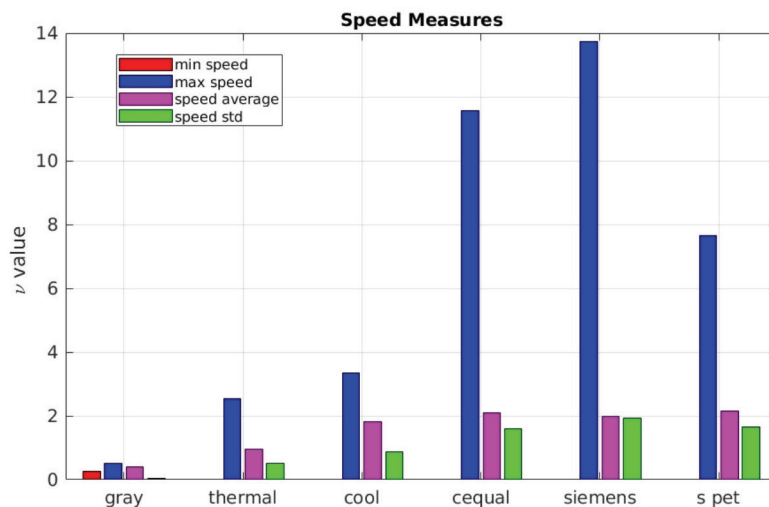


Figure 6. The statistical measures of speed including maximum and minimum values as well as the mean (average) and standard deviation is plotted in the bar chart. As can be seen, those measures are small for "gray" and "thermal", but, for other colormaps, the variation in the speed is remarkable

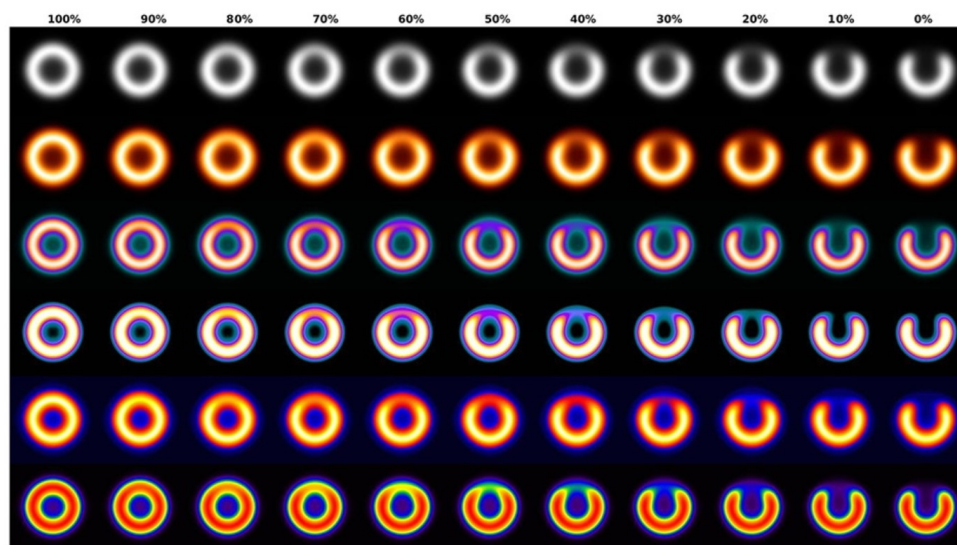


Figure 7. The tomographic slices reconstructed by iterative MLEM method with graded defect severity (100% to 0% myocardial activity). In each row, the images are visualized with one of 6 colormaps. The defect is located on the anterior wall

MLEM: Maximum likelihood expectation maximization

reader's eye are to a high extent subjective, and readers, based on their own experience and perception, decide to estimate the severity of the defect. Although this way of interpretation constitutes the main part of decision making in clinical settings, it seems that approximately noticeable differences exist among different colormaps. In "gray" and "thermal" colormaps, defects seem to the reader's eye to change more slowly than other colormaps, "cool", "cequal" and "s pet" in particular. In other words, the perceptual impact of changes is dramatic, in close agreement with the results shown in Figures 2, 4.

To make the results more objective, the colormaps were analyzed with more quantitative metrics. For this task, certain sample points are selected on the tomographic slice of the cardiac phantom (Figure 8), and the values of those pixels are analyzed in terms of relative count or intensity, lightness or luminance and the ΔE^{76} . Figure 9A, B, C, D, E, F consists of 6 parts according to defect severity from 100% uptake (or no defect) to 50% of maximal uptake. The circumferential profiles of intensity or count, lightness and ΔE^{76} curve are plotted for various levels of defect severity. Thus, the three curves are comparable for each defect severity. In each part, the upper panel displays the tomographic image with its specific defect severity in six colormaps, and in the lower panel, the curves of count, lightness and ΔE^{76} are plotted. In each plot, the profile starts from the inferior wall and then through the septum reaches the anterior wall (which is defective). Subsequently, by passing through the lateral wall reaches a similar point in

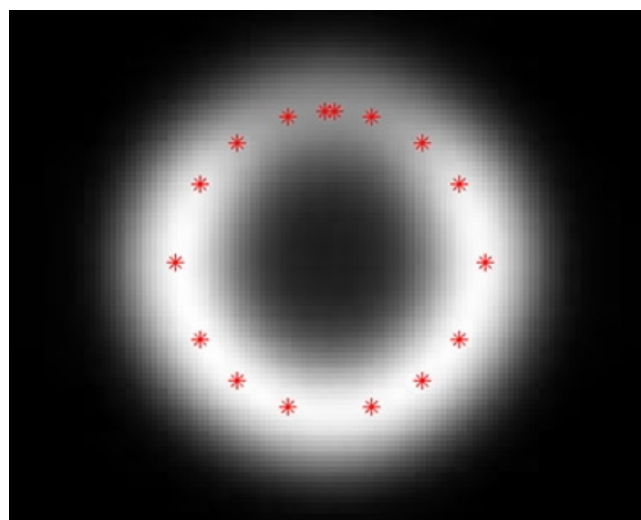


Figure 8. A tomographic slice with a defect on the anterior wall. The sample points are placed on the peak myocardial uptake radially. Sample points are dispersed symmetrically, closer to the anterior wall. It starts from inferior wall going through the septum, then anterior wall (the defect) and through lateral wall it ends in inferior wall to plot the circumferential profile of the short axis slice of the LV

LV: Left ventricle

the inferior wall. Thus, the defect is in the mid-profile. Each plot has three curves. The black curve denotes the true value of the sample points (or counts), and the red and cyan curves indicate lightness (L^*) and ΔE^{76} . When there is

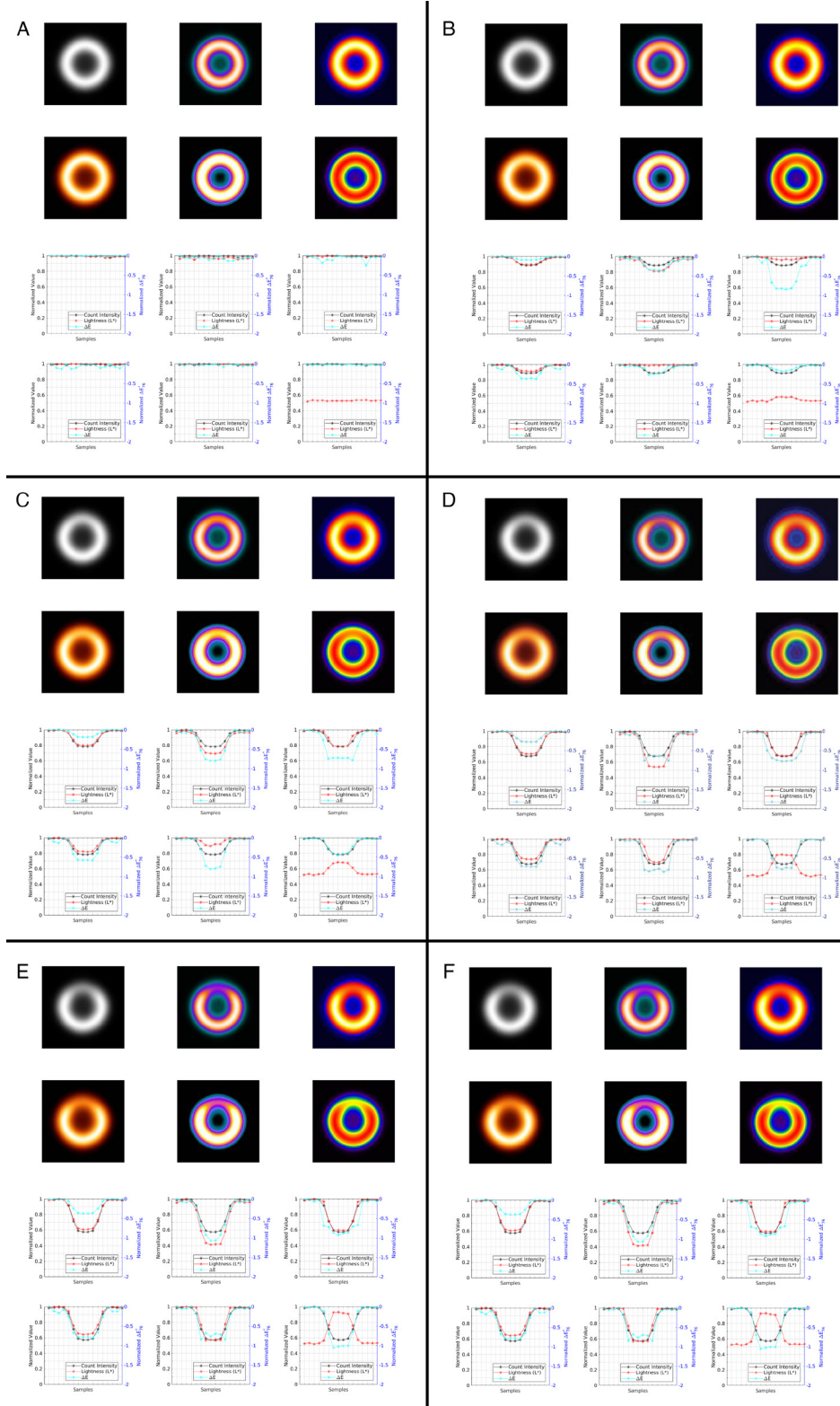


Figure 9. Tomographic slices and circumferential profiles of intensity or count, lightness, and ΔE_{76} are plotted for various levels of defect severity as 0% (A), 10% (B), 20% (C), 30% (D), 40% (E) and 50% (F)

no defect, all the curves are at the highest level (equals 1 or 100%), except for "s pet", in that the lightness is not at its maximum value. In the phantom with 10% defect (or 90% of maximum value), discrepancies among curves appear. In "Siemens" colormap, the ΔE^{76} drops sharply. This pattern is concordant with the way it appears to the eye. In 80% defect, ΔE^{76} curve, in "gray" colormap drops more slowly than other curves of other colormaps. Instead, the defect is more noticeable in "cool", "Siemens" and "cequal". In "s pet" colormap, the lightness curve rises paradoxically although the count intensity and ΔE^{76} curves are matching. In 70% defect, again, the curves are in good agreement in "thermal", "Siemens" and "cequal". However, a consistent lag exists in "gray" colormap. Up to 50% defects, the curves maintain their expected pattern, but in defects more severe than 40%, lightness and ΔE^{76} curves in "cool" and "cequal" rise paradoxically, and in "thermal" they start to slow down in descent. In "Siemens", the falling pattern of the three curves continues. For "s pet" colormap, an erratic pattern of lightness and ΔE^{76} curves exists.

Discussion

Colormaps have long been used for color coding of inherently gray-intensity images. Broadly speaking, they are generally employed for scientific data visualization. The process of pseudo-coloring, apart from its aesthetical purposes, seeks ways to facilitate the conveyance of information qualitatively and semi-quantitatively (1,2,3,4). For this reason, many colormaps have been designed, some of which are frequently used in medical imaging. In nuclear medicine, in particular, a few colormaps including "grayscale", "inverted gray", "thermal", "cool", "Siemens", and traditionally, "rainbow" or "s pet" are of high interest among practitioners in this field (23,24,25). In other words, these colormaps are specific to this particular domain and are mostly used based on convention, convenience, accessibility, etc. Few studies have investigated the effect of using colormaps in clinical images. As mentioned before, the inherently complex and multi-aspect nature of the assessment of colors and the difficulty in devising accurate methods for quantification and modeling preclude much investigation on this topic.

Colors are mathematically modeled using several color spaces and particular measures. Of the several color models, CIELAB, developed by the Commission Internationale de l'Eclairage, is considered the most perceptually uniform. The model comprises a pyramidal geometry. The vertical axis indicates lightness, and the horizontal plane includes a coordinate system with a- and b-axes. The a-axis ranges the colors between red and green on the positive and negative sides, respectively. The b-axis includes colors between blue

and yellow on the negative and positive sides. The colors of points that reside in the origin or on the lightness axis are a shade of gray between black and white. Distances between two colors (or points) are calculated in Euclidean space, as mentioned in the methods section. The main application of this color space is to create perceptually uniform colormaps. However, the related measures are employed as perceptual metrics. ΔE^{76} is the first ever metric introduced in 1976 (4,20,21). This study was intended as a preliminary investigation for characterizing and specifying various colormaps and demonstrating their advantages and pitfalls. In cardiac SPECT and PET images, the colormaps are generally used to categorize the level of uptake or defect severity as well as the changes from one image to another corresponding image (for example, in comparing stress and rest images, termed as reversibility, or in serial imaging) (9,24,26). Therefore, the performance of colormaps in each of these applications can impact the process of decision making. Since there is not much knowledge and understanding about the precise performance of each image, it is expected that the interpretation of such images is highly subject to errors. More prevalently, readers are not fully aware of the advantages and disadvantages and of where to use each one based on the application. Considering all these facts, we intended to analyze their performance quantitatively as much as possible. This warrant consistent objectivity in our comparative evaluation.

There are no perfect colormaps that are applicable to various purposes. Therefore, according to the basic characteristics and specifications of a colormap, the selection should be made. The performance of colormaps is assessed using several metrics, including distance, speed, linearity, uniformity, discriminative power, order, and smoothness (4,20,21). Unfortunately, there is no unanimously accepted nomenclature for these attributes and measures. This issue creates some ambiguities when comparing the colormaps. Despite this fact, we tried to use measures that are more clarified in definition, such as distance and speed. Heatmap Euclidean distance matrix is an easy-to-use method for visualizing pairwise distance between every two colors in the colormap at a glance. Because the distance is based on its perceptual impact, colormaps with softer fluctuations or variations are more desirable. This fact defines the property of uniformity and smoothness measured by local speed (the speed between the colors of two neighboring levels in colormap. Uniformity can be evaluated on the basis of statistical parameters, including standard deviation. According to our findings in Figure 3, "gray" and "thermal" colormaps are perfectly linear. After that, "Siemens", "cool" and "cequal" come. In this respect, "s pet" colormap is far

from a linear property. Now, the question that arises here is which one is important, whole-range or partial-range linearity. The answer, to the best of our knowledge, lies in the application. Since in cardiac SPECT and PET images, the defects with a severity from 0% to 50% are much more of interest, the upper half of the range of the colormaps are of greatest importance (levels from 128 to 256). In semi-quantitative analysis, the grading of defect severity is as follows: mild 10%-25%, moderate 25%-50%, and severe >50% reduction in count compared with maximal myocardial uptake or equivalent semi-quantitative scores of 1, 2, and 3, respectively (26,27). Thus, mild, moderate, and severe lesions are distinguishable. Although "gray", "thermal" and "Siemens" as well as "cool" and "cequal" are all linear in the upper half, the slope of the curves of "cool" and "cequal" is steeper. This leads to higher and more accentuated discrimination or power. In technical terms, "cool" and "cequal" colormaps inherently have an effect similar to that of an exponential function during intensity transformation or a wider dynamic range. This can result in more distinguishing ability. In summary, the whole-range linearity of "gray" and "thermal" does not seem to be an advantage for the analysis of cardiac SPECT and PET images, in contrast to the common belief. It may be of more benefit in other applications because of maintaining whole-range linearity. In the figures, the upper and lower halves are separated by a green vertical line, which indicates 50% of the maximal myocardial uptake. For a more in-depth analysis, the distance is computed using the perceptual metric, ΔE^{76} . The distance, as shown in Figure 4, is calculated based on the color of the maximal value or level 256 in the colormap as a reference. Therefore, the distance between all other levels with level 256 is plotted. Likewise, the output of the upper half in "gray" colormap ranges from 0 to 50 compared to "thermal" colormap, which ranges from 0 to 100, and is steeper in the uppermost sub-range. This finding also provides a more distinguishing ability or discriminative power to categorize lesions as mild or moderate. The results are even better for "cool" and "cequal". In "Siemens", the distance curve is remarkably steep at the rightmost sub-range, enabling the distinction of mild lesions from normal uptake. The findings presented in Figure 5 confirm these results. The speed in "gray" is too low (in the range of 0.2 to 0.5) to distinguish levels from each other. Considering this property, "cool", "cequal" and "thermal" are more favorable compared to other colormaps.

In the next section of this project, the above analysis is performed on simulated cardiac images, which are displayed using the six colormaps. To accomplish this goal, a cardiac phantom with various graded levels of defect severity in

the anterior wall was designed, and the respective color differences between normal walls with defects were qualitatively or visually analyzed (Figure 7). This graded defect severity with its color appearance in various colormaps enables readers to make a consistent qualitative assessment. In addition, the circumferential profile of the myocardial walls in the short-axis section is plotted for true pixel intensity or count, changes in the lightness of colors, and color distance as the main perceptual metric. The profile is drawn from samples with the highest myocardial intensity in the radial direction. The order for plotting is from inferior, septal, anterior, lateral, and then inferior walls, so that, at the defect region, the three curves may overlap or separate from each other congruently or incongruently. Thus, one can evaluate the effect of each colormap in the perception and estimation of defect severity. As expected, some underestimate the defects, which means that the curves for lightness and color difference (ΔE^{76}) may fall behind the intensity or uptake curve. It is worth noting that this lag occurs at different levels as the grade of defect severity increases. In contrast, others may overestimate the defects, which can be interpretable as the curves for lightness and color difference ΔE^{76} rush into the lead or get ahead. As an example (Figure 9), in "gray" colormap, the curve for ΔE^{76} always lags the other two curves. Therefore, there is a consistent underestimation, but the pattern is uniform over the entire range of defect severity. In "thermal" colormap, the curves of ΔE^{76} and lightness and uptake run at the same pace. This pattern is more favorable than that of "gray", but the output intensity accentuation is less than that of other colormaps, which means that it lacks enhanced discrimination between mild, moderate, and severe defects. The "cool" and "cequal" colormaps have a favorable effect on readers. However, in "cequal" colormap, the distinction between normal and mild defects is so small that the curve of lightness reaches a plateau. In "Siemens" colormap, the defect is markedly overestimated at first (a large gap between ΔE^{76} and uptake curve) but remains steady after that. It is interpretable as the distinction between mild and moderate/severe defects becomes vague. Finally, for "s pet" color map, the curves for ΔE^{76} and lightness change irregularly, which hampers the accurate estimation of defect severity. Therefore, its use in the interpretation of cardiac SPECT and PET images is strongly discouraged.

This project focuses on a particular technical aspect of color perception, namely color difference, with quantitative metrics. The process of perception itself is much more complicated and involves neurophysiological and psychophysical aspects. Despite the quantifications based on these methods, human perceptions are different

among individuals. In parallel, the issue can be investigated by observer performance models using receiver operating characteristic curve analysis among different readers or observers. However, all these methods are complementary for modeling and quantitative evaluation of human perception of colors and lesion detection (15,16,28).

Conclusion

There certainly is not a perfect colormap for various purposes of image visualization. Selection depends on the application. Of the 6 colormaps investigated in this study for estimating defect severity, “grayscale” is less favorable than others, and “thermal” performs slightly better. The “s pet” or rainbow, which is traditionally used by many practitioners, is strongly discouraged. The “Siemens” colormap suffers from decreased discriminating power in the range of mild to moderate/severe defects. In contrast, the “cool” and “cequal” colormaps outperform the other colormaps employed in this study to some extent, although they have some shortcomings.

Ethics

Ethics Committee Approval: Not necessary.

Informed Consent: Not necessary.

Financial Disclosure: The author declared that this study received no financial support.

References

1. Silva S, Sousa Santos B, Madeira J. Using color in visualization: A survey. *Comput Graph* 2011;35:320-333.
2. Stone MC. Representing colors as three numbers. *IEEE Comput Graph Appl* 2005;25:78-85.
3. Zhou L, Hansen CD. A Survey of Colormaps in Visualization. *IEEE Trans Vis Comput Graph* 2016;22:2051-2069.
4. Bujack R, Turton TL, Samsel F, Ware C, Rogers DH, Ahrens J. The Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps. *IEEE Trans Vis Comput Graph* 2018;24:923-933.
5. Rogowitz BE, Treinish LA. How not to lie with visualization. *Comput Phys* 1996;10:268-273.
6. Candell-Riera J, Santana-Boado C, Bermejo B, Armadans L, Castell J, Casáns I, Jurado J, Magriñá J, de la Rosa JN. Interhospital observer agreement in interpretation of exercise myocardial Tc-99m tetrofosmin SPECT studies. *J Nucl Cardiol* 2001;8:49-57.
7. Hansen CL. Digital image processing for clinicians, part I: Basics of image formation. *J Nucl Cardiol* 2002;9:343-349.
8. Siennicki J, Kuśmierk J, Kovacevic-Kuśmierk K, Biełkiewicz M, Chiziński K, Płachcińska A. The effect of image translation table on diagnostic efficacy of myocardial perfusion SPECT studies. *Nucl Med Rev Cent East Eur* 2010;13:64-69.
9. Cooke CD, Vansant JP, Krawczynska EG, Faber TL, Garcia EV. Clinical validation of three-dimensional color-modulated displays of myocardial perfusion. *J Nucl Cardiol* 1997;4:108-116.
10. Kamei R, Watanabe Y, Sagiyama K, Isoda T, Togao O, Honda H. Optimal monochromatic color combinations for fusion imaging of FDG-PET and diffusion-weighted MR images. *Ann Nucl Med* 2018;32:437-445.
11. Kimpe T, Tuytschaever T. Increasing the number of gray shades in medical display systems—how much is enough? *J Digit Imaging* 2007;20:422-432.
12. Geissbuehler M, Lasser T. How to display data by color schemes compatible with red-green color perception deficiencies. *Opt Express* 2013;21:9862-9874.
13. Race AM, Bunch J. Optimisation of colour schemes to accurately display mass spectrometry imaging data based on human colour perception. *Anal Bioanal Chem* 2015;407:2047-2054.
14. Uemura M, Asai Y, Yamaguchi M, Fujita H, Shintani Y, Sanada S. Psychophysical evaluation of calibration curve for diagnostic LCD monitor. *Radiat Med* 2006;24:653-658.
15. Krupinski EA. The role of perception in imaging: past and future. *Semin Nucl Med* 2011;41:392-400.
16. Burgess AE. Visual perception studies and observer models in medical imaging. *Semin Nucl Med* 2011;41:419-436.
17. Emery KJ, Volbrecht VJ, Peterzell DH, Webster MA. Variations in normal color vision. VI. Factors underlying individual differences in hue scaling and their implications for models of color appearance. *Vision Res* 2017;141:51-65.
18. Sharma G. Digital color imaging handbook 2017;6:1-797.
19. MacDonald LW. Using color effectively in computer graphics. *IEEE Comput Graph Appl* 1999;19:20-35.
20. Representation. In: Solomon C, Breckon T. Fundamentals of digital image processing: a practical approach with examples in matlab. Oxford: John Wiley & Sons, Ltd 2011;12.
21. Color image processing. In: Gonzalez RC, Woods RE. Digital image processing. 3rd Ed. New Jersey: Pearson Education, Inc. 2008;394-460.
22. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671-675.
23. Travin MI. Pitfalls and Limitations of Radionuclide and Hybrid Cardiac Imaging. *Semin Nucl Med* 2015;45:392-410.
24. Hansen CL. The role of the translation table in cardiac image display. *J Nucl Cardiol* 2006;13:571-575.
25. Peace RA, Lloyd JJ. Comment on “The role of the translation table in cardiac image display”. *J Nucl Cardiol* 2007;14:134; author reply 134-135.
26. Dorbala S, Ananthasubramaniam K, Armstrong IS, Chareonthaitawee P, DePuey EG, Einstein AJ, Gropler RJ, Holly TA, Mahmarian JJ, Park MA, Polk DM, Russell R 3rd, Slomka PJ, Thompson RC, Wells RG. Single Photon Emission Computed Tomography (SPECT) Myocardial Perfusion Imaging Guidelines: Instrumentation, Acquisition, Processing, and Interpretation. *J Nucl Cardiol* 2018;25:1784-1846.
27. Tilkemeier PL, Bourque J, Doukky R, Sanghani R, Weinberg RL. ASNC imaging guidelines for nuclear cardiology procedures: Standardized reporting of nuclear cardiology procedures. *J Nucl Cardiol* 2017;24:2064-2128.
28. Chakraborty DP. New developments in observer performance methodology in medical imaging. *Semin Nucl Med* 2011;41:401-418.