COLORNET: Importance of Color Spaces in Content based Image Retrieval

Judy Gateri[1+], Dr. Richard Rimiru[2++], Dr. Micheal Kimwele[2++]

[*] Department of Computing, Jomo Kenyatta University of Agriculture and Technology

Abstract
The mainstay of current image recovery frameworks is Content-Based Image Retrieval (CBIR). The most distinctive retrieval method involves the submission of an image query, after which the system extracts visual characteristics such as shape, color, and texture from the images. Most of the techniques **use** RGB color space to extract and classify images as it is the default color space of the images **when** those techniques fail to change the color space of the images. To determine the most effective color space for retrieving images, this research discusses the transformation of RGB to different color spaces, feature extraction, and usage of Convolutional Neural Networks for retrieval.
Keywords:
CBIR, RGB, HSV, CIELab, Convolutional Neural Networks

1. Introduction

Content-Based Image Retrieval (CBIR) is considered a significant application in the computer vision field. The extraction of distinguishing characteristics that provide a better description of the images in the database usually follows the process of matching the similarity of the appropriate images. As a result, retrieving images requires comparing previously acquired images that represent the images' important aspects rather than matching them to the whole images. Since this is a critical requirement for the domains, numerous strategies are used to accurately get the appropriate images from the databases. Because the most recent algorithms produced fail to achieve a color space modification to the image, most studies employ RGB color space to identify images [1]. As a result, we propose in this study that we use multiple color spaces to see which one is the most useful.

A color space is a technique method for specifying, creating, and visualizing a color. Defining color involves determining its characteristics such as brightness, colorfulness, and hue while the description of color by a computer involves the determination of the amount of red, green, and blue phosphor emission needed to complement a color. Color spaces are termed as an abstract mathematical model for describing colors as numbers but to a computer, these numbers do not contain any intrinsic meaning [2]. Various color spaces are utilized for various applications. Some applications are attached to specific equipment (device dependent) while others are compatible with whatever equipment they are utilized [3].

1.1 Color Spaces

A color space is a precise operation of colors which involves combining the color profile that is supported by various physical gadgets and supports the reproduction of various representations of the color. A color space can be discretionary for example with actually acknowledged color distributed to a bunch of actual colors samples with the ones corresponding to the assigned ones. A “Color space” is a helpful device/tool that understands the abilities of a specific device and digital record. Reproducing colors in a different device determines whether one can keep the shadow/feature detail, the saturation of the color, and the level of compromisation.

The human perception of color used to differ depending on the processing target used in the application, which is why requirements can't accommodate both. Color spaces refer to fit structures in which values are addressed in accordance with the International Lighting Commission (CIE) standard.
The improvement of the color frameworks affects visual correspondence/communication like TV, broadcasting frameworks, clinical image processing, and processing video signal just as in the field of computers like graphic equipment and printing [4].

1.1.1 RGB color space

The RGB color space is a color space often used to describe images. There are three color components: red, green, and blue. The representation is an 8-bit 24-bit representation associated with each R, G, and B

Manuscript received May 5, 2023
Manuscript revised May 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.5.5

filter. This means that each channel has a value extension in the range 0-255. Figure 1 shows the RGB representation. One notable limitation of the RGB color space is that it is not perceptually uniform, implying that the determined separation in RGB space does not directly correspond to the perceptual color distinction [5]

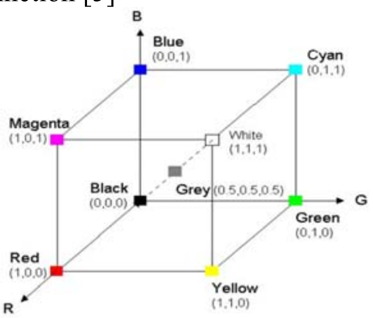

Figure 1: RGB representation.

1.1.2 HSV Color space

HSV stands for Hue, Saturation, and Value and was promoted with the goal of establishing a human color perspective. Describes the color according to the given range. The range is as follows: Hue: 0 to 360, saturation: 0 to 100, value: 0 to 100. HSV is represented as a cylindrical geometry that starts with a red button with a hue and accurate measurements of 0 ° and progresses to the next step. It returns to essential green at 120 °, essential blue at 240 °, and finally red at 360 °. In HSV space, the vertical axis along the focal point represents the intensity of the value, the red axis opposite the intensity axis represents hue, and the distance opposite the point from the intensity axis represents saturation. The figure below shows HSV color space hexacone representation.

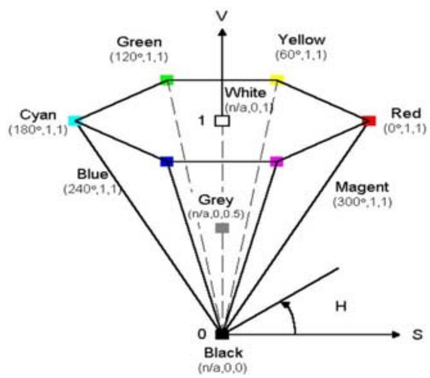

Figure 2: Color hexacone for HSV representation.

1.1.3 LAB Color Space

Commission Internationale de lclairage) (CIE) LAB color space, also known as the CIE L * a * b * or "Lab" color space, was defined by the CIE (International Commission on Illumination) in 1976. The lab color space is often used as an alternative design when monitoring other hardware, so it is device independent. [5] The CIELAB color space conveys colors in three ways. L * represents the brightness from black (0) to white (100), a * represents green () to red (+), and b * represents blue (-) to yellow. (+). CIELAB was developed with the aim of comparing similar measurements of mathematical changes in these properties with common similar measurements of externally perceived changes.
This paper is organized as follows. The next section describes the related work. Section 3 discuss methodology, section 4 experiment, analysis and findings. Finally, in section 5 conclusion.

2. Related Work

Several image retrieval methods that use a mixture of color and texture properties have been proposed. For instance, Raja Proposed a CBIR methodology that involved the use of the Sobel and Canny method in the area of interest of the image as well as the applied HSV color space. A neural network for classification was used where various similarity metrics such as Manhattan, Euclidean, Chebyshev, Hamming, and Jaccard distance were used to estimate the similarity distance. The experiment was performed on Corel-1k and Corel-5k datasets and the results were estimated as 87.33% and 68.93% accuracy and 86.36% and 68.47% precision respectively [7]. CBIR method based on psychophysical and neurobiological characteristics to simulate human visual systems was proposed [8]. The paper showed that human visual characteristics could be presented effectively and more accurately in the HSV color space than in the L*a*b* color space. The results of the experiment indicated Mean Average precision as 55.89% and 56.57% for Lab and HSV respectively. Shreyank N Gowda and Chun Yuan, on the other hand, used a straightforward CNN model comprised of two convolutional layers preceded by a max-pooling layer. Then came a dropout layer, two more convolutional layers, one more max pooling layer, one more dropout layer, and finally a dense layer. The model investigated the significance of color spaces for classifying an image. It was evaluated using different databases including Cifar 10 and Cifar 100 and produced different performances [1].

Some deep learning-based methods for object recognition [9] and image classification using neural networks [10], [11], have been discussed. An all-inclusive review of CBIR systems has also been discussed in [12]. Lastly, Ouhda Mohamed et al used Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) for retrieving images very fast and also extracting and classifying features of the images.

The proposed model used the ImagNet database. The learning of a CNN-SVM comprised the optimization of the coefficients of the models as well as random initialization for minimizing the classification error in the final result. It was concluded that a user needed to understand the coefficients of the convolution kernel for the extraction of the appropriate characteristics and the accuracy in combining these features. The model produced an accuracy of 98.5% [13]

Networks with a large number of layers gave higher accuracy results on ImageNet. By way of illustration, [14] had 19 layers, [15-16] had more than 100 layers. It was deduced as indicated in [17] that the wider you go, the higher the accuracy not the deeper you go.

3. Methodology

The proposed approach involves two phases including extracting the features and then matching the similarity. In the first phase, features are extracted using three major color spaces whereas the second phase is matching similarity using CNN as indicated in figure 3.

3.1 Feature Extraction

The image features are extracted using RGB as a default form, then transformed to HSV and LAB. In the second step, the query image feature set is matched with the dataset feature set using Convolutional Neural Networks.

When it comes to the extraction of the features it involves feeding the properties of the signals instead of the signals themselves to a classifier which can be trained like a Support Vector Machine [18]) which improves classification performance. In particular, they have to use feature extractors that are non-linear, i.e. by using kernels in the context of Support Vector Machine which maps input signal space divisions that are not straightly detachable (linearly) into linearly distinguishable space polarities [19] [20].

The figure below demonstrates how features are extracted and retrieved.

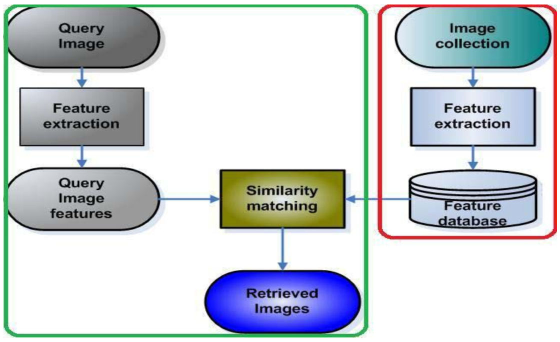

Figure 3: Diagram of the proposed method for image extraction and retrieval

3.1.1 RGB Color space

Because the images in the Cifar 10 dataset are in RGB format, there was no need to transform them.

3.1.2 HSV Color space

The process of converting RGB to HSV is shown below. The R, G, and B values correspond to the red, green, and blue channels, respectively. H is the hue channel, S is the saturation channel, and V is the value channel.

$$R' = R/255,\ G' = G/255,\ B' = B/255,$$
$$Cmax = max(R', G', B'),$$
$$Cmin = min(R', G', B') \quad (1)$$
$$\Delta = Cmax - Cmin$$

Hue calculation is given by:

$$\begin{cases} 60^0 * \left(\dfrac{G' - B'}{\Delta} mod\ 6\right), Cmax = R' \\ 60^0 * \left(\dfrac{B' - R'}{\Delta} + 2\right), Cmax = G' \\ 60^0 * \left(\dfrac{R' - G'}{\Delta} + 4\right), Cmax = B' \end{cases} \quad (2)$$

Saturation calculation is given by:

$$S \begin{cases} 0, & Cmax = 0 \\ \dfrac{\Delta}{Cmax}, & Cmax \neq 0 \end{cases} \quad (3)$$

Value calculation is given by:

$$V = Cmax \quad 4)$$

3.1.3 LAB color space

LAB is another common color space. Where L stands for lightness, "a" stands for green-red element, and "b" stands for blue-yellow. You can change an RGB image to a LAB by first changing the RGB image to an XYZ image.

CIE XYZ was designed to encompass the most remote numerical reaches of human vision. The X, Y, and Z filters are derived from the R, G, and B filters to avoid displaying negative values. Y represents the brightness, Z represents a filter close to the blue filter, and X represents the blend of cone response curves selected to be perpendicular and non-negative to the brightness [21]. The XYZ image is obtained by utilizing RGB.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = A.\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (5)$$

3.2 Matching Similarity

The model utilized for matching similarity is Convolutional Neural Networks. The main pros of CNN compared to its predecessors is its automatic detection of significant features without any human intervention.

CNN is designed for automatic and adaptive learning of the spatial hierarchies of the attributes through backpropagation through the utilization of several main components like convolution layers, pooling layers, and fully connected layers [22].

The network is made up of three convolutional layers, each followed by a max-pooling and a dropout layer. This was followed by a dense layer, a dropout layer, and another dense layer.

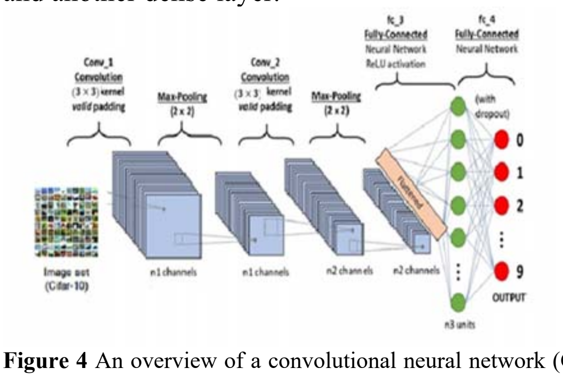

Figure 4 An overview of a convolutional neural network (CNN) architecture and the training process.

3.3 Dataset

CIFAR 10 dataset which is one of the most commonly used datasets for machine learning was used. The CIFAR-10 dataset contained 60,000, 32x32 color images with 10 different classes. The 10 different classes represented airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks.

The dataset was separated into five preparing clusters and one test bunch, each with 1000 images. The test batch had 100 randomly selected images from each class. The training cluster consisted of the remaining images which were randomly ordered. However, some clusters contained a bigger number of images from one class to another. The training batches each contained exactly 5000 images from the dataset, as well as 10 random images.

4. Experimental Research, Analysis, and Findings

Cifar 10 data was used in assessing the performance of the proposed model in this section

Performance Metrics. To evaluate the performance of the proposed methods, we used the Precision and Recall as defined below:

Precision (specificity) determines the capability of the system in retrieving similar images only to the query image

$$precision = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ Number\ of\ images\ retrieved}$$

The Recall involves measuring the system's ability in retrieving all the applicable models, on the other hand, precision involves measuring the system's capability to retrieve only applicable models.

$$Recall = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ Number\ of\ relevant\ images}$$

Precision and recall computation is dependent on the number of query images (from the test dataset) and the image similarity as retrieved from the image database. In this scenario, the quantity of important things recovered is the quantity of the returned images that are like the inquiry image. The overall number of things recovered is the quantity number of images that are returned by the model.

F1-score - The harmonic mean of recall/sensitivity and precision [26] is used to calculate F1-score

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision}$$

The harmonic mean discourages classifiers that sacrifice one measure for another too drastically. Among the four standard measures, precision involves assessing the correctness of the classification, while the others involve the evaluation of the correctness regarding each class [27].

ALGORITHM

- A database of 32×32 normalized images was created.
- RGB format image is inputted.
- Color transformation is done.
- As input features to CNN's first layer, various image features from various color spaces are used.
- Compute image database using the same techniques.
- Use the CNN model to query image features from the image database.
  - o To search image:-
  - o While not converge Do
  - o For each image Do
  - o Map them in non-linear space using an activation function (ReLU). (ReLU - f(x) = max (0, x))
  - o To learn model parameters, continue forward pass and backpropagation. (∂xℓ(i−a)(j−b)∂yℓ−1ij=ωab∂x(i−a)(j−b)ℓ ∂yijℓ−1=ωab)
  - o End for
  - o End while

- A Similar image is retrieved from the image database.
- Calculate and print the percentage similarity of the image which is r retrieved.
- Compute Precision, Recall, F1score, Receiver Operator Characteristic (ROC), and Area Under Curve (AUC).

The experiment was carried out with RGB, HSV, and CIELab at 50, 100, 150, and 200 Epochs, respectively. Tables 1 and 2 show the accuracy, mean average precision (MAP), and mean average recall (MAR): -

Table 1: Summary of a Confusion matrix for different color spaces under 50 and100 Epochs

	50 Epochs			100 Epochs		
	RGB	HSV	LAB	RGB	HSV	LAB
Accuracy %	73	69	69	74	71	71
MAP %	73	70	69	74	74	72
MAR %	73	69	69	74	71	71
Fscore %	73	69	69	74	71	71

Table 2: Summary of a Confusion matrix for different color space under 150 & 200 Epochs.

	150 Epochs			200 Epochs		
	RGB	HSV	LAB	RGB	HSV	LAB
Accuracy %	75	73	72	71	71	70
MAP %	75	73	73	71	70	71
MAR %	75	73	72	71	71	70
Fscore %	75	73	72	71	71	70

Figures 5 and 6 represent the graph model loss and accuracy for RGB color space after 150 epochs.

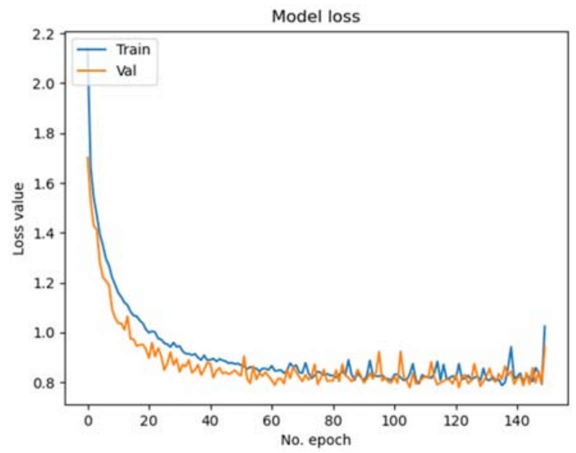

Figure 5: Model loss for RGB at 150 epochs

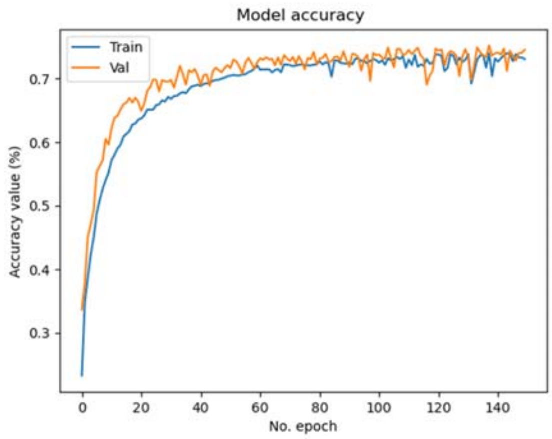

Figure 6: Model Accuracy for RGB at 150 epochs

Figure 7 shows the Receiver Operator Characteristic (ROC) and Area Under Curve (AUC) of RGB color space.

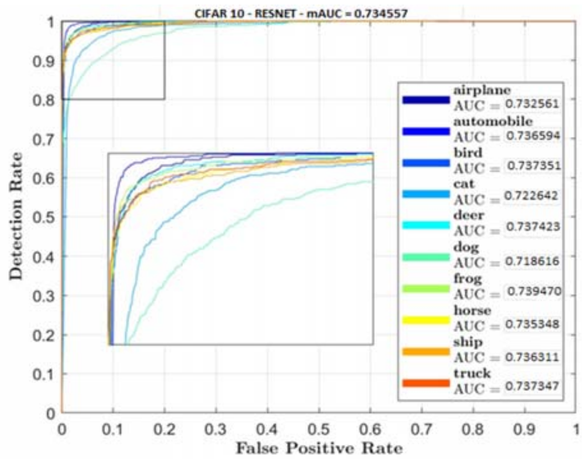

Figure 7: ROC and AUC of RGB color space.

The study showed more effective retrieval when more epochs were used. The study started with 50 epochs and the researcher kept on increasing by the range of 50. The study further showed an increase in retrieval in 100 epochs and 150 epochs respectively. There was however a drop in retrieval when the study used 200 epochs and therefore the study concluded the optimal epoch retrieval was 150 epochs.

Additionally, the study closely examined the confusion matrix to exactly examine each class under different color spaces. According to the study, airplanes, automobiles, frogs, horses, ships, and trucks were well represented using RGB, with accuracy, mean average precision (MAP), and mean average recall (MAR) exceeding 73%. The Cat class was also poorly represented in the study, with 51%, despite Accuracy, Mean Average Precision (MAP), and Mean Average Recall (MAR) being 73 percent under 50 Epochs.

Although the accuracy did not vary significantly across the different color spaces, closer examination with the aid of confusion matrices revealed that some classes were highly recognized despite the system's overall accuracy, while others were poorly recognized. The experiment presented that airplanes, Automobiles, frogs, Ships, and Trucks were well recognized using RGB color space.

The experiment also showed that Deer and Dog were poorly recognized with RGB color space but were well recognized using HSV. The experiment further showed Cat was well recognized using LAB. Although their precision and recall rate percentage was low compared to that of the system.

5. Conclusion

According to the study, preprocessing images in different color spaces produces different results. Even though the accuracy does not differ significantly on its own, we discovered that there was no 100 percent correlation between the outcomes after further investigation using a confusion matrix. We found that specific classes are far better represented in specific color spaces. The research also found some classes like the truck, ship, and automobile are well represented using RGB. We also noticed birds, dogs, and deer were well represented using HSV. We also found Cat was well represented using LAB.

The study concluded that it is critical to transform images into different color spaces for better and more efficient retrieval. The research also concluded the importance of combining color space with either shape or texture to increase the recognition rate.

Acknowledgment

The completion of this undertaking could not have been possible without participation of many people whose names may not been enumerated and therefore we would like to express our deep appreciation to them.

References

[1] Gowda SN, Yuan C. ColorNet: Investigating the importance of color spaces for image classification. In Asian Conference on Computer Vision 2018 Dec 2 (pp. 581-596). Springer, Cham.

[2] Mokrzycki WS, Tatol M. Color differenceΔ E-A survey. Machine graphics and vision. 2011 Apr;20(4):383-411.

[3] https://en.wikipedia.org/wiki/Adobe_RGB_color_space

[4] Al-saleem, R. M., Al-Hilali, B. M., & Abboud, I. K. (2020). Mathematical Representation of Color Spaces and Its Role in Communication Systems. Journal of Applied Mathematics, 2020.

[5] van den Broek EL, Kok T, Schouten TE, Vuurpijl LG. Human-centered content-based image retrieval. InHuman Vision and Electronic Imaging XIII 2008 Feb 14 (Vol. 6806, p. 68061L). International Society for Optics and Photonics.

[6] Al-saleem RM, Al-Hilali BM, Abboud IK. Mathematical Representation of Color Spaces and Its Role in Communication Systems. Journal of Applied Mathematics. 2020 May 18;2020.

[7] Raja R, Kumar S, Mahmood MR. Color Object Detection Based Image Retrieval Using ROI Segmentation with Multi-Feature Method. Wireless Personal Communications. 2020 Jan 1:1-24.

[8] Qazanfari H, Hassanpour H, Qazanfari K. Content-Based Image Retrieval Using HSV Color Space Features. International Journal of Computer and Information Engineering. 2019 Sep 5;13(10):537-45.

[9] Meng X, An Y, He J, Zhuo Z, Wu H, Gao X. Similar image retrieval only using one image. Optik. 2016 Jan 1;127(1):141-4.

[10] Xu YY. Multiple-instance learning based decision neural networks for image retrieval and classification. Neurocomputing. 2016 Jan 1;171:826-36.

[11] Zarchi MS, Monadjemi A, Jamshidi K. A semantic model for general purpose content-based image retrieval systems. Computers & Electrical Engineering. 2014 Oct 1;40(7):2062-71.

[12] Ajam A, Forghani M, AlyanNezhadi MM, Qazanfari H, Amiri Z. Content-based Image Retrieval Using Color Difference Histogram in Image Textures. In2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS) 2019 Dec 18 (pp. 1-6). IEEE.

[13] Mohamed, O., Mohammed, O., & Brahim, A. (2017, October). Content-based image retrieval using convolutional neural networks. In First International Conference on Real Time Intelligent Systems (pp. 463-476). Springer, Cham.

[14] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015 Dec 1;115(3):211-52.

[15] Srivastava RK, Greff K, Schmidhuber J. Training very deep networks. InAdvances in neural information processing systems 2015 (pp. 2377-2385).

[16] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[17] Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint arXiv:1605.07146. 2016 May 23.

[18] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[19] C. M. Bishop, Pattern recognition and machine learning. Springer, 2009.

[20] Wiatowski, T., & Bölcskei, H. (2017). A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Transactions on Information Theory, 64(3), 1845-1866.

[21] https://www.image-engineering.de/library/technotes/958-how-to-convert-between-srgb-and-ciexyz

[22] Yamashita R, Nishio M, Do RK, Togashi K. Convolutional neural networks: an overview and application in radiology.

[23] Insights into imaging. 2018 Aug 1;9(4):611-29.

[24] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," vol. 1502, 2015, https://arxiv.org/abs/1502.03167.

[25] Si, L., Xiong, X., Wang, Z., & Tan, C. (2020). A Deep Convolutional Neural Network Model for Intelligent Discrimination between Coal and Rocks in Coal Mining Face.

[26] G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. http://arxiv.org/abs/1207.0580,2012

[27] Cios KJ, Pedrycz W, Swiniarski RW. Data mining methods for knowledge discovery. Springer Science & Business Media; 2012 Dec 6.

[28] Suh JH. Comparing writing style feature-based classification methods for estimating user reputations in social media. SpringerPlus. 2016 Dec 1;5(1):261.

Judy Gateri is currently a candidate for the degree of Doctor in Philosophy (PhD) Computer Science at the Jomo Kenyatta University of Agriculture and Technology and acquired Msc. Software Engineering from the same university. She serves as a full-time faculty at the Jomo Kenyatta University of Agriculture and Technology. Her research works revolve around Machine Learning and computer vision systems.

Richard Rimiru received his Doctorate degree in Computer Science and Technology from Central South University (CSU), Changsha, Hunan, P. R. China) in 2013. His MSc (Distinction) in Computer Science, Awarded in 2002 from National University of Science and Technology (NUST), Bulawayo, Zimbabwe) in 1998. BSc. (First Class Honours) in Statistics and Computer Science, Awarded in 1999 (Jomo Kenyatta University of Agriculture and Technology). His research interests include Bio-inspired computing – Neural nets, evolutionary computing, Artificial immune systems, Image processing – feature extraction and manipulation methods

Michael W. Kimwele is a lecturer in the Department of Computing, Jomo Kenyatta University of Agriculture and technology (JKUAT). He holds a BSc. Mathematics and Computer Science-First Class Honours from JKUAT (2002), a Masters in Information Technology Management from University of Sunderland UK (2006) and a Doctorate in Information Technology from JKUAT (2012). His research interests include Information systems management, Information Technology Security, Electronic Commerce, Mobile Computing, Social implications of computer applications, Human Computer Interaction, and Computer Ethics.