International Journal of Computational Intelligence and Applications
Vol. 22 (2023) 2350018 (16 pages)

DOI: 10.1142/S1469026823500189

Fake Colorized Image Detection Based on Special Image Representation and Transfer Learning

Khalid A. Salman*, Khalid Shaker† and Sufyan Al-Janabi‡

College of Computer Sciences and Information Technology
University of Anbar, Iraq
**khalidtmeme@yahoo.com*
†khalidalhity@gmail.com
‡sufyan.aljanabi@uoanbar.edu.iq

Received 19 June 2022
Revised 4 September 2022
Accepted 16 February 2023
Published 1 April 2023

Nowadays, images have become one of the most popular forms of communication as image editing tools have evolved. Image manipulation, particularly image colorization, has become easier, making it harder to differentiate between fake colorized images and actual images. Furthermore, the RGB space is no longer considered to be the best option for color-based detection techniques due to the high correlation between channels and its blending of luminance and chrominance information. This paper proposes a new approach for fake colorized image detection based on a novel image representation created by combining color information from three separate color spaces (HSV, Lab, and Ycbcr) and selecting the most different channels from each color space to reconstruct the image. Features from the proposed image representation are extracted based on transfer learning using the pre-trained CNNs ResNet50 model. The Support Vector Machine (SVM) approach has been used for classification purposes due to its high ability for generalization. Our experiments indicate that our proposed models outperform other state-of-the-art fake colorized image detection methods in several aspects.

Keywords: Image colorization; color spaces; CNNs; transfer learning; SVM.

1. Introduction

Colorization is a recent image editing technology that colorizes grayscale images with realistic colors, and potentially it may use to recolor an original colored image for fun or illegal purposes. Specifically, in the forensic field, digital content such as images and video can become digital evidence within a legal process, in which this type of data helps to confirm the facts under examination. Thus, if the digital proof has been intentionally manipulated, it can directly affect the course of the investigation. In colorization, the original objects in the image are not buried, obscured, or replaced; only the color of things is changed.[1]

Colorization methods are classified as scribbling methods, example methods, and fully automatic approaches. Scribble methods are examples of supervised algorithms

*Corresponding author.

that depend on users to start assigning colors for grayscale picture pixels. The milestone work[2] suggested that the nearby pixels with comparable intensities should have the same colors. Several more ways were developed consequently, such as that described in Ref. 3, which creates texture and color dictionaries using sparse representation and appropriately adds colors to the photos. Scribble-based techniques need substantial human effort in terms of both time and expertise.[4]

The example-based methods[5,6] often need users to oversee the system by supplying a color picture to be a reference comparable to the grayscale image. The colors of the reference picture are then transferred to the target grayscale picture by looking for similar objects or patterns. The effectiveness of these approaches is determined by the quality of the reference image.

Unlike the previously-automatic supervised methods, fully automatic modes do not need supervision when applying the colorization operation. The proposed approach in Ref. 7 used a deep neural architecture that has been trained to include semantically relevant characteristics of different complexity into colorization. The work in Ref. 8 developed a colorization system based on deep learning that uses CNNs trained on more than a million color photos in the Lab (Lab) color space. Using the lightness L, the proposed approach can predict the image's conformable (a and b) channels in the Lab color space. Another work[9] presented an image colorized system that uses both local and global prior's image characteristics to colorize grayscale photos automatically. The deep network based on CNN allows for the efficient combination of local information based on small picture patches and global priors calculated utilizing the full picture. The superior performance of fully automatic colorization methods can be noticed in Fig. 1. In this paper, we focus on how to identify the fake colorized images that are made by these techniques.

The main contributions of this research are focused on the following topics:

- A novel image representation cbaH is created by combining color information from three separate color spaces and selecting the most different channels from each color space to reconstruct the new image representation.

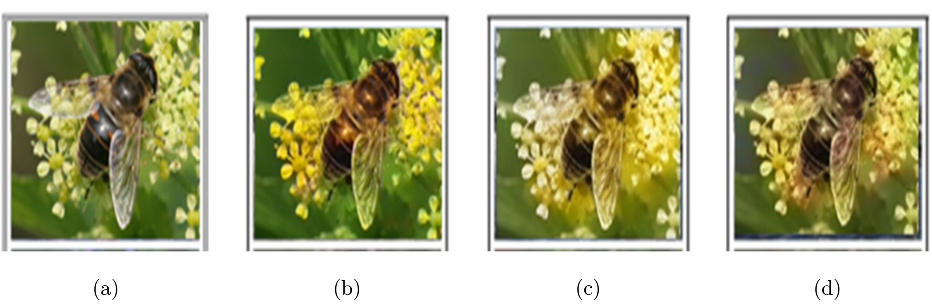


Fig. 1. (a) Real images, (b), (c) and (d) are colorized images created by colorization methods suggested in Refs. 7–9, respectively.[10]

- A transfer-learning-based ResNet50 model is proposed for feature extraction from input images.
- Detailed results are provided for the transfer-learning ResNet50 model related to classifier performance using two different classifiers, SVM and Neural Networks NNs.

This paper is arranged as follows. Section 2 presents related works of fake colorized image detection. Section 3 explains our proposed fake colorized image detection system. Next, Sec. 4 provides the experiments dataset and results with a comparison with the state of art methods. Finally, Sec. 5 provides a brief conclusion for the work.

2. Related Works

The first considered and studied fake colorized image problem was Guo *et al.*[4] They suggested two successful detection approaches based on hand-crafted statistical features for fake colorized image detection, Histogram-based (FCID-HIST) and Feature Encoding-based (FCID-FE). FCID-HIST uses the most distinguishing bins to generate four features for detection: dark channel, hue, saturation, and bright channel features, which are intended to detect forgeries. In FCID-FE models, data are sampled using the Gaussian mixture model (GMM), and fisher vectors are created to use statistical differences better. Both approaches employ SVM as a classifier. The experimental results demonstrate that there is still opportunity for improvement in detecting performance. Pilli *et al.*[11] used hand-crafted features (hue, saturation, dark channel, bright channel, and an alpha channel) with the Gaussian kernel model. To enhance the model, a fuzzy classification model was employed. This method performs better by (4%) compared to FCID-HIST and (2.5%) compared to FCID_FE that used in Ref. 4. Li *et al.*[12] used a neural network with hand-crafted features of normalized histograms computed for the R, G, B, h, s, and v channels and employed the cosine similarity of normalized histogram distributions. Experiments demonstrate that the proposed method significantly outperforms FCID-FE and FCID-HIST detection methods proposed in Ref. 4.

The work of Agarwal *et al.*[13] is proposed to use Local Binary Pattern (LAB) as a feature extractor method on real pictures and fake pictures. From top to bottom and left to right, LAB was applied over the entire image. As a feature vector, a histogram of the final image is employed. For binary classification, Linear Discriminant Analysis (LDA) has been employed as a classifier. The proposed approach provided a percentage error value about (3.47%). Yu *et al.*[14] presented a strategy that uses Lateral Chromatic Aberration LCA with Histogram characteristics (LCAH) for detection. The proposed approach was based on the idea of that the Recolored Images (RIs) have a lower number of LCA properties than Natural Images (NIs). A five-dimensional vector has been used to train SVM for classification purpose. Experiments demonstrate that the proposed method performs better than the methods proposed in Ref. 4.

Zhuo *et al.*[15] proposed a deep-learning model for fake colorized image detection based on channel-wise convolution. The deep-learning framework used was WISERNet (Wider Separate Then-Reunion Network); it takes an RGB picture as an input and performs channel-wise convolution on the picture's R, G, and B channels, respectively. The first convolutional layer corresponds to the proposed network's separate stage. The second convolutional layer, called "reunion," forms input by concatenating the three separate sets of the output channels. The rest of the proposed network consists of a series of convolutional and fully connected layers until it reaches the final layer, consisting of two neurons representing "fake" and "real" predictions. The system performance boost on top of specific handcrafted features is used in Ref. 4. Ulloa *et al.*[1] used a transfer learning model with a custom model consist of parallelism with two convolutional layers blocks to extract features and classify purposes. The proposed models provided good results of Half Total Error Rate HTER with custom model (9%) and with transfer learning model (2.6%). The works of Yan *et al.*[16] proposed a deep learning model that takes the original picture with two derived from it, inter-channel correlation and illumination consistency as inputs, and output the probability of the image is recolored or not. The proposed approach has accuracy about (86.89%). Jijina *et al.*[17] also used same two derived inputs in Ref. 16 based on illumination consistency and inter-channel correlation of the original input image into CNNs for feature extraction and classification. The proposed approach has validation accuracy about (82.46%). Quan *et al.*[18] used an end-to-end CNNs framework and aimed to increase the system's generalization capability by inserting a negative sample into CNNs. The results showed that both accuracy and generalization were improved compared with methods in Ref. 4.

In our proposed approach, a deep learning technique based on transfer learning has been used to reduce the computational cost. Transfer learning uses the knowledge gained during the solution of one problem to another problem. A special proposed image representation has been used as an input for the ResNet50 pre-trained model to extract useful features. The extracted features are then used to train SVM to detect real and fake colorized images.

3. The Proposed Approach

The architecture of the proposed approach is shown in Fig. 2. The main aims of our system are to improve the accuracy and generalization and reduce the training time of the model. To fulfill these objects, the system is divided into three stages. The first is the preprocessing stage that converts RGB images to a new representation to make fake and real images more differentiable, increasing detection accuracy. The second stage uses transfer learning based on the pre-trained deep learning model ResNet50 for feature extraction; using transfer learning models reduces the computational requirement in terms of time and system components. At the final stage, deep residual features (2048-d) extracted from pre-trained ResNet50 are used to train SVM for classification purposes.

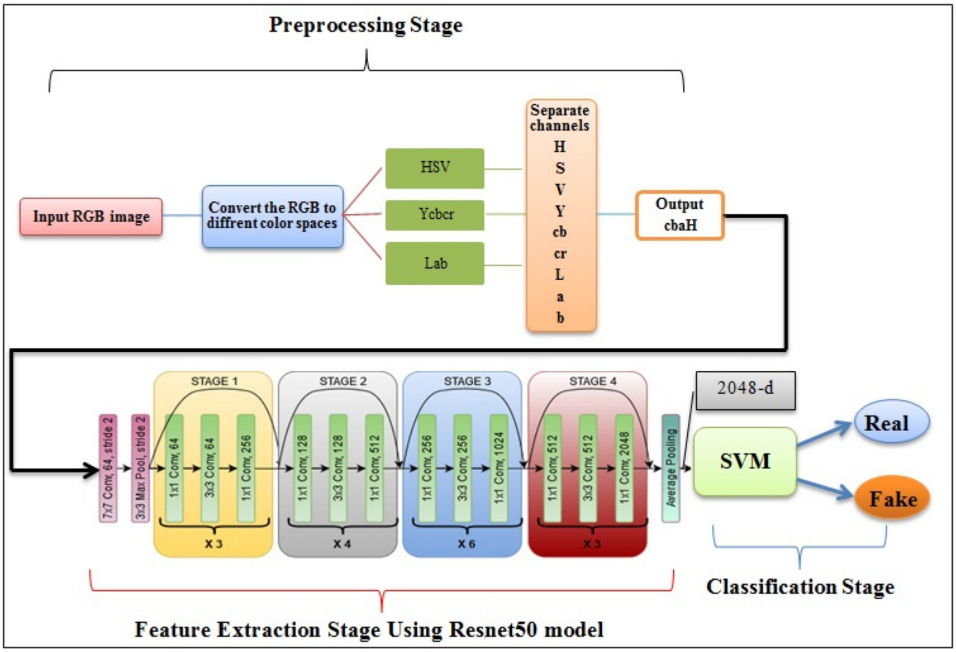


Fig. 2. The architecture of the proposed fake colorized image detection system.

SVM has been widely used in classification due to its robust theoretical base, high generalizability, minimal sensitivity to the curse of dimensionality, and capacity to identify global classification solutions. An explanation of each stage is provided in the next sub-sections.

3.1. *Preprocessing stage*

The core point of any image forgery detection is the image itself; from this idea, and to obtain better image representation containing useful information for our problem that deals with colors. A new image representation has been proposed based on merging color information from different color spaces.

3.1.1. *Image color spaces*

Numerous color spaces present color information that makes certain calculations easier or provides a more intuitive method for identifying colors. For instance, the RGB color space defines color as the proportions of red, green, and blue colors. Other color spaces separate the chrominance and luminance information. In this work, RGB image has been converted into different color spaces and generated a new image representation by mixing color information from different color spaces. Three color spaces, HSV, Lab, and Ycbcr, have been tested to find the best image representation.

Each one of these color spaces consists of three channels. The three-color spaces are explained below.

A. *HSV color space*

HSV, which stands for hue, saturation, and value, is a color system that closely correlates to human color perception. This is why it is so common when a user is requested to select colors. The following describes the significance of the three components[19]:

- Hue (H): The real color property that defines red, green, etc.
- Saturation (S): The proportion of white color combined with the color. A color with more white will have a low saturation value.
- Value (V): Relates to the luminance of a color.

The transformation from RGB to HSV space is given as follows:

$$V = \max(R, G, B), \tag{1}$$

$$S = \begin{cases} \dfrac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} & \text{if } \max(R,G,B) \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

$$H = \begin{cases} \text{undefined} & \text{if } S = 0, \\ \dfrac{G - B}{\max(R,G,B) - \min(R,G,B)} & \text{if } R = \max(R,G,B), \\ 2 + \dfrac{B - R}{\max(R,G,B) - \min(R,G,B)} & \text{if } G = \max(R,G,B) \\ 4 + \dfrac{R - G}{\max(R,G,B) - \min(R,G,B)} & \text{if } B = \max(R,G,B). \end{cases} \tag{3}$$

B. *Ycbcr color space*

The Ycbcr color space is denoted by one luminance component, Y, and cb with cr represents chrominance components. The Y component of this color space represents the intensity of the light. The cb and cr components represent the relative strengths of the blue and red components concerning the green component. Ycbcr color space simulates human vision, i.e., it capitalizes on the properties of the human eye. The equation describing the conversion between the RGB color system and Ycbcr color space is as follows[20]:

$$\begin{bmatrix} Y \\ cb \\ cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.279 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{4}$$

C. *CIE Lab color space*

The CIE Lab format is derived from the CIE XYZ format and is considered perceptually uniform, which means that colors that have the same distance in Lab space have the same perceptual distance, which is achieved using a uniform chromaticity scale. The L parameter of a Lab color is represented with values ranging from (0 to 100), while a and $b$are the red/blue and yellow/blue chromaticity values ranging from (-128 to 127). Conversion of RGB input image into Lab space has a two-step process: First, the input RGB image is converted into CIE XYZ color space. Then CIE XYZ is converted into CIE Lab. The transformation is carried out according to the following formula[21]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.35758 & 0.180423 \\ 0.212671 & 0.71516 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{5}$$

$$L = \left\{ \begin{matrix} 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & \text{if } \frac{Y}{Y_n} > 0.008856 \\ 903.3\left(\frac{Y}{Y_n}\right) & \text{if } \frac{Y}{Y_n} \leq 0.008856 \end{matrix} \right\}, \tag{6}$$

$$a = 500 * \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right), \tag{7}$$

$$b = 200 * \left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right), \tag{8}$$

$$\text{where} \quad f(n) = \left\{ \begin{matrix} n^{\frac{1}{3}} & \text{if } n > 0.008856 \\ 7.787 * n + \frac{16}{116} & \text{if } n \leq 0.008856 \end{matrix} \right\}, \tag{9}$$

where the luminance ($Y_n = 1.0$), and the chrominance ($X_n = 0.950455$, $Z_n = 1.088753$) for the D65 white point.

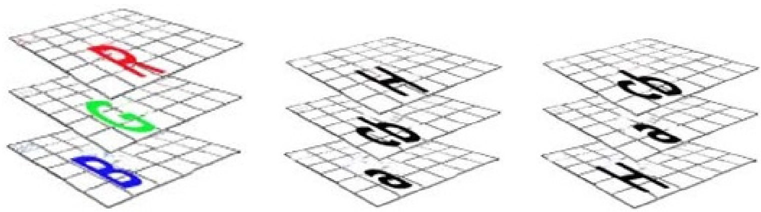
3.1.2. *Image representation*

In a fake colorized image detection problem, the color information (chromaticity) is more important than luminance (lightness). For that reason, and to provide a global view and diversity of color information, the representation of the input image has been changed to increase the ability to detect and differentiate between original and fake colorized images. The input image that is already in RGB has been converted to the three suggested color spaces and separated all nine channels (H, S, V, L, a, b, Y, cb, and cr) to select the best three channels (one from each color space) and reuse them to build the new image representation.

To detect the most different channels, (2000) original images from ImageNet dataset[22] with their corresponding fake colored pictures made by colorization techniques in Refs. 7–9 have been compared to check the similarity using Mean Square Error MSE. The best three channels with the low similarity between real and fake images have been selected to represent the image. Table 1 shows the similarity degree (mean of MSE for 2000 images from each colorization method has been calculated) of different image channels from different colorization methods.

Table 1. Similarity degree (higher has lower similarity) of image channels with different colorization method.

Image Channels	Colorization Method[7]	Colorization Method[8]	Colorization Method[9]
H	**0.1078**	**0.1270**	**0.1209**
S	0.0668	0.0721	0.0641
V	0.0056	0.0055	0.0055
L	0.0011	0.00027	0.00026
a	**0.0604**	**0.0555**	**0.0548**
b	0.0568	0.0516	0.0495
Y	0.00070	0.00016	0.000015
cb	**0.0024**	**0.0034**	**0.0029**
cr	0.0023	0.0032	0.0029



(a) (b) (c)

Fig. 3. Image channel concatenation: (a) standard RGB image, (b) Hcba, (c) cbaH.

From Table 1, it is clear that the (H, a, and cb) channels have lower similarity values in their color spaces, making them more suitable to differentiate between original and fake images. As shown in Fig. 3, the concatenation of channels also affects the nature of features extracted from the image, especially with pre-trained CNNs models (i.e., order Hacb provide different features of Hcba). For that reason, all concatenations possible have been tested and found that cbaH was the order that provided the best results.

3.2. *Features extraction based on transfer learning*

Transfer learning is an approach in which a model generated for one job is utilized as the foundation for a model on a different challenge. The idea is to transfer the weights learned by a network at (task one) to a new (task two). Instead of starting from scratch, the learning process starts with patterns found while doing a task related to the one being learned. Transfer learning is mostly used in tasks that require a lot of computing power, like computer vision, natural language processing, and image classification.[23]

Our model proposed a pre-trained deep residual network ResNet-50 for image feature extraction. ResNet-50 is a deep residual network introduced by He *et al.*[24] and won the ImageNet competition in 2015. The "50" refers to the number of layers it has. The main innovation of ResNet and the main reason we proposed this model instead of any other pre-trained models is the skip connection. The skip connection permits input shortcuts to pass through the block without traversing the other weight layers. This property reduces the computational cost when the shortcut input is passed without multiplying by a layer's weight matrix.[25] ResNet-50 trained on over a million photos from the ImageNet dataset and can classify images into 1000 object categories, including keyboard, mouse, pencil, and numerous animals. Consequently, the network has acquired rich feature representations for a broad number of images. Figure 4 shows the ResNet-50 architecture.[26]

In this work, as illustrated in Fig. 2, the ResNet-50 model has been used for features extraction based on the following steps:

(1) Remove the fully connected layers from the pre-trained ResNet-50 trained on the ImageNet dataset and have an output for different 1000 classes.
(2) Freeze the rest of the ResNet-50 layers to be used as a feature extractor on the new data.
(3) Train SVM classifier using the Deep Residual Features (2048-d) extracted from the new dataset containing original and fake colorized images to classify the output for two classes, real and fake images.

3.3. *Classification stage using SVM*

A support vector machine is a supervised learning technology commonly used in data analysis and pattern recognition for classification and regression analysis. Methods vary based on the classifier's structure and properties. The most well-known SVM is a linear classifier that predicts which of two potential categories each input belongs to. A more accurate description is that a support vector machine creates a hyper plane or group of hyper planes to categorize all inputs in a high-dimensional or unlimited space. The values that are nearest to the classification margin are called support vectors. The goal of the SVM is to maximize the distance between the hyper plane and the support vectors. SVM has been widely used in classification because of its strong theoretical background, high generalization ability, limited sensitivity to

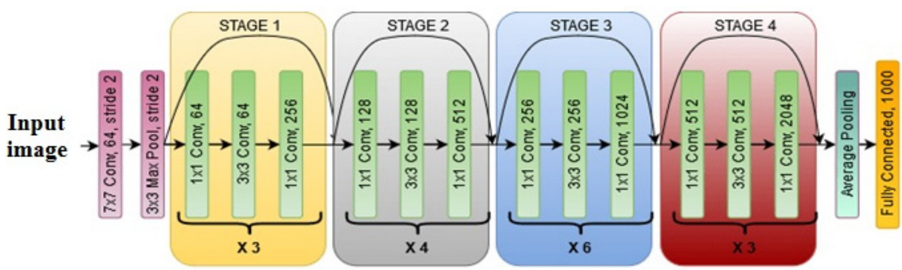


Fig. 4. ResNet-50 architecture.[26]

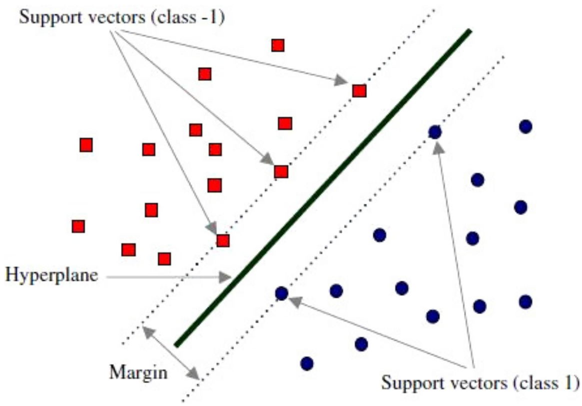


Fig. 5. Example of SVM classification.[28]

the curse of dimensionality, and capacity to identify global classification solutions.[27] Figure 5 shows SVM classification example.

In the proposed approach, features extracted using pre-trained ResNet-50 have been used to train SVM for classification purposes. In fake colorized image detection problems, the original and fake colorized images may produce some sharing features because of their similarity that made us in front of linearly inseparable data. For that reason, the linear kernel function is not recommended in this kind of problem. To solve these issues, Radial Basis Function RBF has been used as a kernel function for SVM that provide the best results with nonlinear data. In RBF, two key hyper-parameters affect the classification process: Gamma and C. The Gamma parameter indicates the extent of a single training example's impact, with low values indicating "far" and large values indicating "near." The C parameter is a regularization parameter used to adjust the model's tolerance for misclassifying data points to reduce generalization error. To find a good fit for hyperparameters, Bayesian optimization has been used to check a list of hyperparameters and find the best values. By minimizing the model's objective function, Bayesian optimization identifies the best collection of hyperparameters for a particular model. This optimization technique picks new hyperparameters carefully at each iteration and often reaches the optimal set of hyperparameters faster than a standard grid search.[29]

4. Dataset and Experimental Results

To evaluate our proposed approach performance and compare the results with other researchers, the widest dataset ImageNet[22] has been used. The dataset is divided into the same format as other related works for comparison purposes. D1 includes 10000 natural true-color photos chosen at random from ImageNet as well as their equivalent fake colorized images produced using methods shown in Refs. 7–9. D1 contains various natural photos, including animals, humans, furniture, and outdoor scenes.

Aside from D1, other datasets have been created to evaluate the effectiveness of the fake colorization image detection systems against various colorization techniques. The database D2 comprises 2000 natural pictures, and their equivalent fake images produced using the colorization method in Ref. 7. The database D3 comprises 2000 real images and their equivalent fake colorized images using the colorization method in Ref. 8. The colorization method in Ref. 9 was used to create the database D4, which comprises 2000 natural photos and their matching created fake images. It is worth noting that the natural photos in D2–D4 do not overlap with each other. The datasets D2–D4 were divided into 1000 pair images (original, fake) for training and the other 1000 pair for testing. For performance evaluation, Half Total Error Rate HTER has been used; it represents the average misclassification rates of positive and negative classes and computes as follows[15]:

$$\mathrm{HTER} = \frac{\mathrm{FP}/(\mathrm{TN}\ +\ \mathrm{FP})\ +\ \mathrm{FN}/(\mathrm{TP}\ +\ \mathrm{FN})}{2}. \tag{10}$$

In our problem, TP corresponds to original images correctly classified, TN corresponds to colorized images correctly classified, FN corresponds to colorized images classified as original images, and FP corresponds to original images classified as colorized images. The experiment conditions are described in Table 2.

At first, the best three channels with the low similarity between real and fake images (H, cb and a) as shown in Table 1 have been selected to represent the image instead of RGB representation. The effect of channel concatenation order on the results was tested to detect the best image representation. Figure 6 shows the result of this comparison using 2000 images of the D2 dataset. The results show that the cbaH channel concatenation provides the best results than other concatenation orders.

From another side, to verify the effectiveness of our proposed image representation, Fig. 7 shows results comparison of standards image color spaces (RGB, Lab, HSV and Ycbcr) against our proposed image representation cbaH using the D2 dataset in term of HTER. Our proposed image representation provides better detection results than standard image color spaces.

To highlight the advantage of SVM in generalization capability, an experiment has been conducted against Neural Networks NNs classifier. The input images are converted from RGB to our proposed representation cbaH. The deep residual features (2048 per image) that extracted using the ResNet50 model from the dataset

Table 2. Experiment conditions of our proposed approach.

System	Fake colorized images detection
Pre-processing	New image representation cbaH
Front-ends	CNNs (ResNets-50)
Back-end	SVM
Number of features per image	2048
Programming Language	MATLAB R2020b
Platform	HP Pavilion dv6 Intel core i7 2.20 GHz, RAM 6 GB

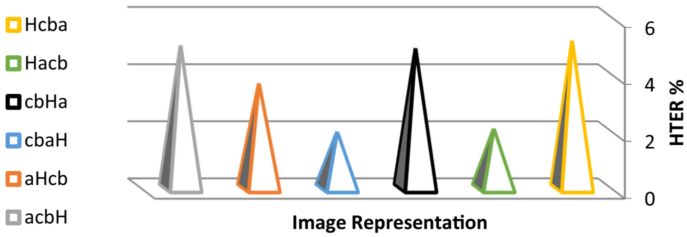


Fig. 6. Comparison of different channels concatenation in terms of HTER in % (lower is better).

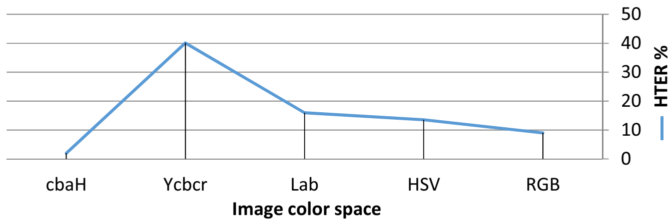


Fig. 7. Comparison of different image color spaces in terms of HTER in % (lower is better).

D1-D4 (containing original and fake colorized images) have been used to train SVM and NNs classifiers. For SVM classifier, the RBF kernel has been used to serves as kernel function. To find the hyperparameters that minimize five-fold cross-validation loss, Bayesian optimization algorithm has been used by the penalty factor C and slack variable gamma to discover the C and gamma parameters with the maximum accuracy in the training model. The function evaluation of SVM trained on D1 dataset is illustrated in Fig. 8, which explains the value of the objective function for each function evaluation.

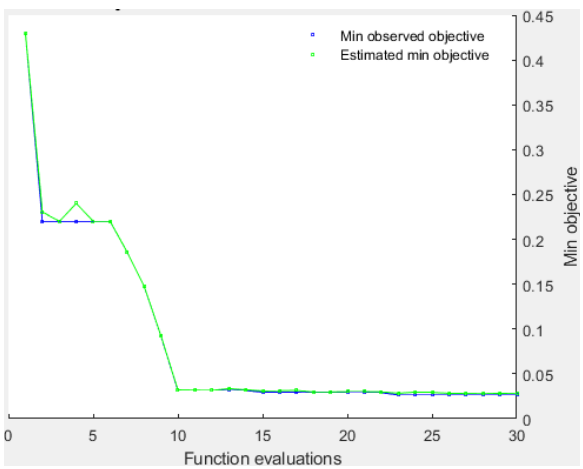


Fig. 8. Minimum objective versus number of function evaluations.

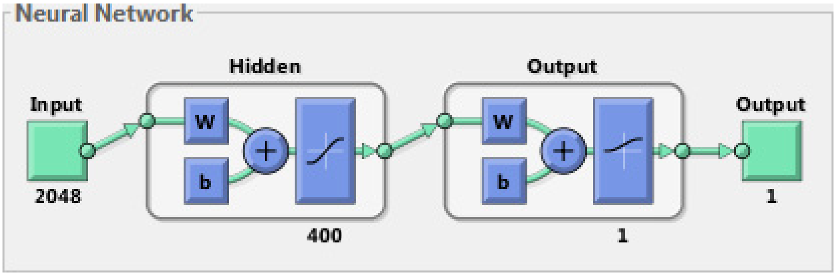


Fig. 9. Architecture of neural network classifier.

For NNs classifier, as shown in Fig. 9, the input features 2048 per image are passed into a two-layer feed-forward network, with sigmoid hidden layer with (400) neuron and softmax output neuron that predict two outputs (fake colorized or real), the network was trained with scaled conjugate gradient back propagation. The performance of the network was calculated using cross entropy, minimizing cross-entropy results in good classification. Table 3 shows a comparison of the performance of SVM against the NNs classifier; results show that SVM outperforms NN in both accuracy and generalization using D2–D4 datasets.

We have experimentally compared our method's classification accuracy and generalization performance with the state-of-the-art methods. The testing results are shown in Table 4. From these results, it's clear that our proposed approach provided fully better results for both accuracy and generalization from the state-of-the-art methods in Ref. 4, 12, and 14, with enhancement for generalization results for the method in Ref. 15.

The researchers in Ref. 1 also depended on transfer learning to detect fake colorized images using (VGG-16) model. They used manual colorization (Photoshop) and automatic colorization method in Ref. 7. We compared our work with their work with the same colorization method and more images used in their work (they used 9506 images, and we used D1 dataset that contain 10000 images). The results show that our method provides better HTER (1.96%) against (2.6%) and (9%) for their transfer learning model and custom model, respectively.

Table 3. Comparison of the performance (HTER, in %, lower is better) of SVM against the NN classifier. Bold represents best accuracy results, and italic represents generalization performance.

	Training Set D2			Training Set D3			Training Set D4		
	Testing Sets			Testing Sets			Testing Sets		
Classifier	D2	D3	D4	D2	D3	D4	D2	D3	D4
SVM	**1.99**	***16.32***	***2.29***	***4.71***	**2.84**	***1.78***	***2.79***	***8.59***	**1.2**
Neutral Network	3.19	*35.71*	*5.27*	*12.40*	3.56	*5.09*	*6.58*	*24.66*	1.99

Table 4. Comparison of the performance (HTER, in %, lower is better) of our method with the state-of-the-art methods. Bold represents best accuracy results, and italic represents generalization performance. The sign (−) means the results do not exist.

	Training Set D2			Training Set D3			Training Set D4		
	Testing Sets			Testing Sets			Testing Sets		
Methods	D2	D3	D4	D2	D3	D4	D2	D3	D4
Guo *et al.* FCID-HIST[4]	22.50	*28.00*	*33.95*	*26.95*	24.45	*41.85*	*38.15*	*43.55*	22.35
Guo *et al.* FCID-FE[4]	22.30	*23.65*	*31.70*	*25.10*	22.85	*34.25*	*38.50*	*36.15*	17.30
Li *et al.*[12]	13.85	*30.45*	*27.00*	*25.80*	12.35	*20.55*	*25.45*	*20.95*	13.85
Yu *et al.*[14]	10.10	*23.5*	*12.95*	—	23.15	—	—	—	14.51
Zhuo *et al.*[15]	**0.65**	***12.75***	*5.65*	*6.2*	**1.6**	*1.9*	*5.2*	***3.7***	**1**
Our Proposed Approach	1.99	*16.32*	***2.29***	***4.71***	2.84	***1.78***	***2.79***	*8.59*	1.2

5. Conclusion

This work aims to identify the fake colorized images from original images. In this paper, a novel image representation has been proposed to improve fake colorized image detection system. Three different color spaces HSV, Lab and Ycbcr have been used for selecting the most useful channels and it was found that cbaH was the best image representation that made the original and fake colorized images more differentiable. The deep learning technique based on transfer learning of (the ResNet 50) model has been used to reduce the computational cost and extract features from the input image. SVM has been used for classification purposes. The experimental finding indicates that such a transfer learning-based approach is well-suited for fake colorized image detection; it can obtain superior performance than the current baseline methods such as FCID-FE, FCID-HIST, LCAH, WISERNet, VGG-16-based model and custom model. For future works, we propose to evaluate other pre-trained models such as MobileNet or GoogleNet instead of our ResNet 50 model.

References

1. C. Ulloa, D. M. Ballesteros and D. Renza, Video forensics: Identifying colorized images using deep learning, *Appl. Sci.* **11** (2021) 476.
2. A. Levin, D. Lischinski and Y. Weiss, Colorization using optimization, *ACM Trans. Graph.* **23**(3) (2004) 689–694.
3. J. Pang, C. Au, K. Tang and Y. Guo, Image colorization using sparse representation, ICASSP, (2013).
4. Y. Guo, X. Cao, S. Member and I. W. Zhang, Fake colorized image detection, *IEEE Trans. Inf. Forensics Secur.* **6013** (2018) 1–13.
5. G. Charpiat, M. Hofmann and B. Sclolkopf, Automatic image colorization via multimodal predictions. In *Computer Vision–ECCV 2008: 10th European Conf. Computer Vision*, Marseille, France, October 12–18, 2008, Proceedings, Part III 10, pp. 126–139. Springer, Berlin Heidelberg.
6. X. Chen, J. Li, D. Zou and Q. Zhao, Learn sparse dictionaries for edit propagation, *IEEE Trans. Image Process.* **25**(4) (2016) 1688–1698.

7. G. Larsson, M. Maire and G. Shakhnarovich, Learning representations for automatic colorization, in *Proc. Fourteenth European Conf. Computer Vision*, Part IV (2016), pp. 577–593.
8. R. Zhang, P. Isola and A. Efros, Colorful image colorization, in *Proc. Fourteenth European Conf. Computer Vision*, Part III, pp. 649–666 (2016).
9. S. Iizuka, E. Simo-Serra and H. Ishikawa, Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification, *ACM Trans. Graph.* **35**(4) (2016) article no. 110.
10. Yuanfang Guo (Andy), Yuanfang Guo's personal website, http://irip.buaa.edu.cn/andyguo/index.html [accessed on 28 September 2021].
11. N. Pillai, A. Kanthe and S. Bhattacharjee, Enriching fake colorized image detection, *IJRAR - Int. J. Res. Anal. Rev.* **6**(2) (2019) 255–261.
12. Y. Li, Y. Zhang, L. Lu, Y. Jia and J. Liu, Using neural networks for fake colorized image detection, in *Advances in Digital Forensics XV, IFIP AICT 569* (2019), pp. 201–215.
13. S. Agarwal, D. Sharma and K. Jung, Forensic analysis of colorized grayscale images using local binary pattern, in *2020 10th Int. Conf. Cloud Computing, Data Science & Engineering (Confluence)* (2020), pp. 507–510.
14. Y. Yu, N. Zheng, T. Qiao, M. Xu and J. Wu, Distinguishing between natural and recolored images via lateral chromatic aberration, *J. Vis. Commun. Image Represent.* **80** (2021) 103295.
15. L. Zhuo, S. Tan, J. Zeng and B. Li, Fake colorized image detection with channel-wise convolution based deep-learning framework, *APSIPA Annual Summit and Conf. 2018* (2018), pp. 733–736.
16. Y. Yan, W. Ren and X. Cao, Recolored image detection via a deep discriminative model, *IEEE Trans. Inf. Forensics Secure.* **14** (2019) 5–17.
17. T. Jijina, L. Koshy and G. Warrier, Detection of recoloring and copy-move forgery in digital images, *2020 Fifth Int. Conf. Research in Computational Intelligence and Communication Networks (ICRCICN)*, November 26–27, 2020, CHRIST (Deemed to be University), Kengeri Campus, Bangalore, India.
18. W. Quan, D.-M. Yan, K. Wang, X. Zhang and D. Pellerin, Detecting colorized images via convolutional neural networks: Toward high accuracy and good generalization, (2019), pp. 1–13, http://arxiv.org/abs/1902.06222.
19. E. Wetzer and H. Lohninger, Image processing using color space models for forensic fiber detection, *IFAC-PapersOnLine* **51**(2) (2018) 445–450.
20. P. Ganesan, V. Rajini, B. Sathish, V. Kalist and S. Basha, Satellite image segmentation based on YCbCr color space, *Indian J. Sci. Technol.* **8** (2015) 35, doi: 10.17485/ijst/2015/v8i1/51281.
21. M. A. Hassan, Perceived quality assessment of color images, *Int. J. Eng. Res. Technol.* **4**(5) (2015) 1592–1595.
22. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and F.-F. Li, ImageNet: A large-scale hierarchical image database, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2009), pp. 248–255.
23. J. Brownlee, A gentle introduction to transfer learning for deep learning, https://machinelearningmastery.com/transfer-learning-for-deep-learning/ [accessed on 10 January 2022].
24. He. Kaiming, Z. Xiangyu, R. Shaoqing and S. Jian, Deep residual learning for image recognition, *IEEE 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA (2016).

25. P. Dwivedi, Understanding and coding a ResNet in Keras, https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33 [accessed on 11 January 2022].
26. Overview of Residual Neural Network (ResNet), https://open-instruction.com/dl-algorithms/overview-of-residual-neural-network-resnet/ [accessed on 12 January 2022].
27. R. Gove and J. Faytong, Machine learning and event-based software testing: Classifiers for identifying infeasible GUI event sequences, in *Advances in Computers*, Chapter 4, Vol. 8 (2012).
28. TechVidvan's R tutorial series, SVM in R for data classification using e1071 package, https://techvidvan.com/tutorials/svm-in-r/ [accessed on 16 January 2022].
29. E. Elgeldawi, A. Sayed, A. R. Galal and A. M. Zaki, Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis, *Informatics* **8**(4) (2021) 1–21.