International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I7P103

Volume 72 Issue 7, 24-36, July 2024

Original Article

The Application of U-Net and Image Colorfulness Frame for Image Colorization Issue

Hani Q. R. Al-Zoubi

Department of Computer Engineering, Faculty of Engineering, Mutah University, Karak, Jordan.

Corresponding Author : Hanirash@mutah.edu.jo

Received: 25 February 2024 Revised: 15 May 2024 Accepted: 14 June 2024 Published: 26 July 2024

Abstract - *This paper proposes a novel approach to the image colorization problem by fusing U-Net architecture with Image Colorfulness Frame (ICF). The proposed method aims to overcome the limitations of existing colorization methods, which often produce pixelated and subpar photos. Color maps are produced using the U-Net architecture using the grayscale input photographs. The color maps' vibrancy is then evaluated using the ICF. The ICF helps to make colorization results more accurate and aesthetically pleasant. Experimental results show that the proposed method outperforms the state-of-the-art methods in terms of colorization accuracy and visual quality. The suggested method has many applications, including colorizing films and photos, restoring black and white images, and colorizing historical images.*

Keywords - *U-Net Grayscale photos, Color maps, Image colorization, Picture colorfulness assessment, Realistic colorization, historical image colorization and black and white image repair.*

1. Introduction

The technique of adding color components to a grayscale image, known as colorization, has garnered significant attention in graphics and computer vision research, especially with the advent of deep learning techniques. A grayscale image inherently contains less information than a color image, and thus, colorizing it can enhance the understanding and interpretation of the visual data. Despite the advancements in this field, the colorization process remains complex due to factors such as the limited amount of data, the diversity of images in the training set, and the accessibility of computational resources [1].

Previously, the technique of applying colour to black-and-white images has involved considerable human effort and a complicated code. However, the appearance of deep learning and artificial intelligence has improved this mission, making it more feasible and automated. It is generally recognized that an abundance of data and a significant amount of training time are required for creating a colorization algorithm from the outset. Current enhancements indicated that it is feasible to efficiently improve these models with limited datasets by employing the most recently developed methods in deep learning [2].

Deep Neural Networks (DNNs) demonstrated incredible effectiveness in overcoming complicated issues in image processing, categorising, and recognition of speech. DNNs provide a fully automated method to image colorization, representing an important advancement over previous methodologies that depended significantly on substantial manual intervention. The convenience of obtaining substantial training data further improves the viability of employing DNN for colorization [3].

Convolutional Neural Networks (CNNs) play a crucial role in an extensive range of image-processing applications that include colorization, categorization, and tagging. In recent years, CNNs have become more common in creating practical image colorization solutions. The U-Net, an instance of CNN architecture initially developed for the segmentation of the image, has been modified to be used for the colorization process. The justification for this is that the process of colourizing images contains an essential similarity with segmentation. Both processes contain the division of an image into recognizable fragments and the following processing of each segment in an appropriate manner [4].

The novelty of this research deceptions in tacking the shortcomings of current colorization approaches by utilising a U-Net architecture. The U-Net's neural network architecture provides efficient image segmentation, an essential process for attaining precise colorization. The U-Net model tackles the problem of information bottlenecks that are typically observed in encoder-decoder models. It accomplishes this by promoting the smooth flow of low-level information across the network. Consequently, the U-Net model is able to efficiently maintain image details during the process of colorization [5].

Fig. 1 Some examples of image coloring

This study improves the field by proposing a method that colorizes grayscale images efficiently and quickly by utilising a small dataset, which contrasts with previous methods that required widespread data and computational resources. The proposed method not only enhances color accuracy but also ensures the colors appear natural to the human eye, thereby improving the overall quality and applicability of automated colorization [6].

1.1. Motivation

Grayscale image coloring by hand requires a lot of time. To avoid all of these efforts and random coloring techniques, a deep learning model was created utilizing the U-net architecture [3]. There are several suggested methods for black-and-white coloring photos. Because the encoder-decoder has an information bottleneck when low-level information flows via the network, the U-Net architecture is employed for coloring. In order to mitigate this problem, the features obtained from the contracting path are additionally linked to the network's up-sampling output layer. Since natural representation channels are supplied into the model as input and the target channels are recovered, more information may be collected from grayscale images, which can then be automatically colored with greater accuracy.

1.2. Problem Statement

Because there is no one right way to conduct colorization, it is a task with many possible outcomes [1]. A U-net predicts the color image from the grayscale image. A wide range of sectors, often color restoration and re-design of historical photographs, are significantly impacted by the coloring of the grayscale image. The final output and input image will have the same dimension, making the colorization challenge both challenging and intriguing [5]. It is, therefore, suitable for image coloring. The major components of the U-Net are encoding bridges and decoding. The latent space of the input is a compact representation of the input image created by the encoding part. When decoding an image, up-sampling and convolution operations are used to recreate the input image the same size as the input images [6]. The bridge's job is to link the encoding and decoding units together. The decoding component of the process concatenates the low-level detail features from the encoding part with the equivalent high-level features.

List of three issues with the model and make recommendations for improvement.

1. It is first crucial to employ a neural network model that has been modified specifically from the classification encoder in order to expect more consistent colorization results for each object. For example, vegetation is never allowed to be blue and must always be green or golden [9].
2. Secondly, a neural network with categorization and regression capabilities or a recommended colorization model can be used. Recent studies indicate that the proposed model can only surpass the baseline or regression model by improving the colour intensity of each object [10].
3. The third requirement is that each model must utilise only one image or dataset, such as photos of human-made objects or landscapes. Data-driven deep-learning colorization frequently results in inconsistent color results for a range of picture types. Landscape photos are one group that, despite the dataset's variety of picture types, displays more consistent, appealing, and accurate color than other types [12].

1.3. Objectives

The main objective of this research study, coloring a grayscale image, requires the use of U-net. The deep Convolutional Neural Network (CNN), known as the U-net, consists of up-sampling and down-sampling channels. They are linked via a bottleneck and skip connection, allowing the feature maps to be copied and concatenated from the encoder to the decoder. A good illustration of this is the Colorful Image Colorization model. It processes an image using a convolutional neural network before applying class rebalancing to produce vivid results. The network's last layer, which reweights each pixel's loss under the rarity of its colors, performs class rebalancing. As part of the network's up-sampling output layer, features from the contracting path are also connected to obtain more information from the grayscale image, allowing for more accurate and realistic automatic coloring [9]. This research aims to improve the U-net network by effectively extracting the characteristics of individuals from newly generated datasets. It is possible to predict the probability distribution of potential colors and human interaction, which allows a range of plausible coloring effects to be obtained in real-time.

$$D.L * a * b \ versus \ RGB \qquad (1)$$

As you may already be aware, when images are imported, they receive a rank-3 (height, width, and colour) broad, with the colour data for the image being contained in the last dimension. The RGB color space represents color in this data, and three integers represent each pixel's red, green, and blue values. As you can see in the accompanying image, the blue color is present in the leftmost image, the "primary image," which causes the blue channel to have greater values and become darker.

In L*a*b color space, each pixel has three values, but these values have different interpretations. Because the first integer (channel), L, encodes the Lightness of each pixel, it shows as a black-and-white image when viewing this channel (the second image in the row below). The *a and *b channels, respectively, are used to represent the quantity of green, red, yellow, and blue light present in each pixel. In the image below, each band of the L*a*b hue can be seen separately.

Over the past several years, various approaches have been used to colorize photos using deep learning. The vehicle in the picture can appear in a variety of distinct and acceptable hues. As a result, it is not applicable to guarantee any hue for it. Another piece, however, approached the issue as a regression task (with some extra changes!). This paper, Colorful Image Colorization, also considered the problem's uncertainty. Each tactic has advantages and disadvantages, but we'll employ a different one in this post. Assume that x represents a grayscale picture, z represents the generator's input noise, and y represents the desired generator output in two channels (it can also mean the two-color channels of an actual image). D is the discriminator, while G is the generating model. The conditional GAN will, therefore, suffer the following loss:

$$L\_cGAN\ (G,D) = E\_(x,y)\ [log\ \ D(x,y)\ ] + E\_(x,z)\ [log\ ⟦(1 - D(x,G(x,z)))⟧\ ] \qquad (2)$$

As the condition to introduce to players in this game, x is provided for both models. Contrary to what you think, it can't be feeding the generator an "n" dimensional random noise vector. Instead, the noise will be included in the generator design in dropout layers (you'll learn something interesting about this in the article's final part). The model will still learn to colorize the images if apply L1 loss, but it will be cautious. When in doubt about which color is best, it frequently settles on hues like "gray" or "brown." To use typical hues, the L1 loss is minimized as much as possible. (identical to the blurring impact of L1 or L2 loss in a super-resolution task). Since the L1 loss (or mean squared error) has a smaller effect on producing grey pictures, it is favored over the L2 loss (or mean squared error).

2. Literature review

The topic of deep learning-based image colorization has seen a few creative approaches that have greatly increased the efficacy and efficiency of the process. Accordingly, the field of image colorization has seen significant advancements with the integration of deep learning techniques, particularly Convolutional Neural Networks (CNNs). Conventional methods require extensive manual involvement. However, recent methods have employed deep learning to fully automate the process, leading to enhanced effectiveness and adaptability [7].

A CNN-based method has been constructed to generate a map feature that preserves both the material content of the grayscale image and the style of the reference image. This method ensures precise colour matching for particular spots by iteratively developing a "random noise map" to recognize the style feature map of the desired area. Researchers have employed colour image segmentation and image fusion methods to improve regional colouring. Deep learning provides the capability of attaining more precise colour assignments by employing a reference image to separate photographs into background and foreground colours.

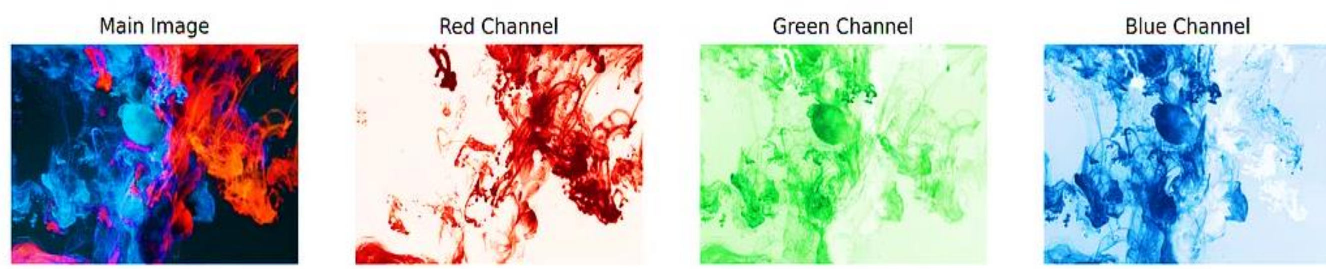

Fig. 2 Red, Green, and blue channels of an image

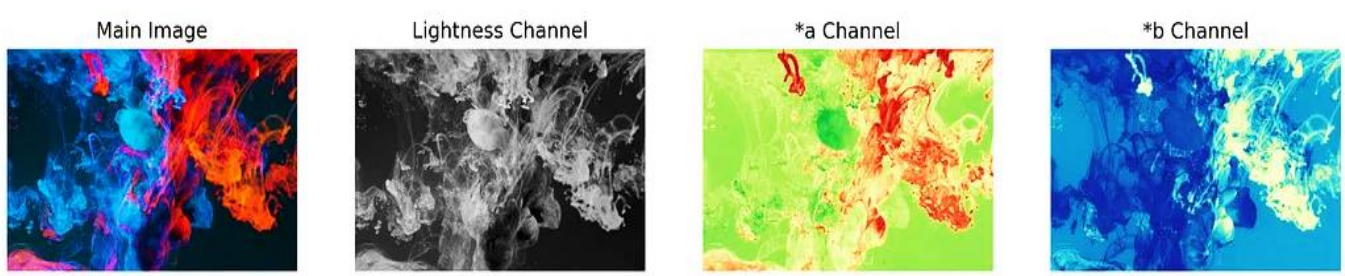

Fig. 3 Lightness, *a, and *b channels of Lab color space for an image.

CNNs play an important role in this method by stripping semantic information from all parts of the image and integrating appropriate colours according to a reference image. Image fusion technique and colour image segmentation are essential methods in this field. Scientists have proposed employing deep learning to enhance the technique of mixed regional colouring, which includes splitting an image into foreground and background colours with the support of a reference image [8].

This method enables more precise colour assignments by recognizing different regions of the background and objects within images. Convolutional Neural Networks (CNNs) are essential in this technique since they obtain semantic data from each part of the image. This information is subsequently employed to allocate appropriate colours according to a reference image. A novel technique has been developed, employing a CNN-based method to generate feature maps that efficiently capture either the content of the grayscale image or the style of the reference image. This technique employs an iterative process that produces a "random noise map" for the purpose of getting the style feature map of the particular region of concern. This ensures a precise colour matching with the desired areas [9]. The model obtains enough information by painting some sections to fully colour the image Figure 2.

The study of translation from image to image has significantly benefited from research carried out in this field. The researchers develop frameworks for interpreting images between two domains, such as colour and grayscale. This technique recognizes coloured and grayscale images as separate domains and facilitates the conversion between them, presenting a powerful and precise method of handling colour assignments. Studies the image-to-image translation have also made a significant contribution to this field. The purpose of these studies is to provide structures for translating images between two domains, such as grayscale and color images. This concept has been used to the colorization problem, enabling more flexibility and precision color assignments [10].

Fig. 4 Technique for coloured images based on indications

Hint-based techniques, such as the integration of scribbles into the image, allow users to direct the colorization process. However, still demanding some human monitoring, this strategy considerably reduces the time and effort required in comparison to conventional methods. Hint-based techniques have been established to simplify the colorization process by employing human inputs, such as scribbles. Farella et al. (2022) proposed a technique that maintains adjacent pixels with the same intensities are allocated the same colour by considering the colorization process as an optimisation problem. By the use of this interactive method, users can effectively produce the desired colour output by through making modifications to their scribbles. Despite these improvements, reproducing representing colours remains challenging. Grayscale images provide sufficient data to identify acceptable colours, but accurate results for atypical or historical topics may require the knowledge of exerts. Automated systems have to achieve a balance between generating visually attractive results and maintaining historical reliability.

Farella et al. (2022) proposed a fascinating hint-based approach in which the colorization process is directed by integrating scribbles into the image. This method approaches the colorization issue as an optimisation problem, ensuring that adjacent pixels with comparable intensities are allocated identical colours. The user has the capability to add or change scribbles interactively in order to achieve the desired colour output. This allows for quick feedback and flexibility [11]. Shihab et al. (2022) developed a significant method called colorization by example. This technique merges user-provided reference photos with automated techniques to simplify the colorization task. The reference image is separated into segments with fixed colours, and an algorithm for supervised classification provides colours to the grayscale image according to these segments. This technique employs Levin's colorization model, in which pixels with high probability estimates are considered "micro-scribbles," so providing a more precise and less time-consuming colorization processing [12]. Despite these developments, there are still difficulties in guaranteeing that the automated systems can effectively recreate colours. Grayscale images often provide sufficient data to estimate acceptable colours, even if the precise original colours are unclear.

For instance, typical things such as grass and sky can be coloured according to their conventional colours in other images. However, when it involves distinctive products or vintage items, the algorithm could have to rely on educated predictions in order to achieve precise outcomes [12]. Despite current methods having made significant improvements in automating the colorization procedure, they frequently involve vast datasets and considerable computational resources. In addition, methods such as hint-based techniques still require substantial human input, which can be time-consuming and require accurate colour selection.

Fig. 5 Image coloring by example technique

This research is novel as it emphasises creating an extremely efficient colorization technique that tackles these constraints through the use of a U-Net construction. The U-Net, initially designed for image segmentation, is highly appropriate for colorization tasks due to its effective preserves low-level information and ensures the smooth distribution of data throughout the network.

This paper introduces a technique that effectively applies colour to black-and-white images by utilizing a limited dataset. It utilises cutting-edge developments in deep learning to generate precise and realistic colours without requiring user participation or a huge amount of data. This technique not only improves the accuracy of colours but also ensures that the colours seem natural to the human eye, which improves the overall quality and applicability of automated colorization procedures.

3. Methodology

Schematic representation of the segmentation and outline selection step based on U-Net. To address the domain bias problem, first separate the noise from an optionally filtered input picture. Entropy masking is used after product classification to decide the end input. Contours are then found and kept in the final picture as far as possible to remove noise. For final picture categorization, the cropped image is then sent to colorization [12].

It requires a lot of time and effort to color a picture, and the quality of the color used greatly influences the outcome. Therefore, the subject of automated picture coloring has significant scholarly ramifications in addition to being useful in real-world situations. As a result of advancements in computer hardware, deep learning technology has successfully attained satisfactory outcomes in the domain of automated colouring [14-17]. This study categorises automatic colouring methods into three classes based on the source of the colour information:

- Coloring pages that are interactive
- Based on previous information
- Based on reference pictures

The majority of users can have their demands satisfied by the coloring approach. However, there are drawbacks, including the inability of users to color several objects in an image using separate reference graphs. According to literary, it is recommended to envision two main applications of CNN in image segmentation and colorization:

1. The model can be trained specifically for a specific animation and used on manga that has not yet been animated to automate the coloring process. In this manner, the model can learn how specific people, animals, and objects should be colored.
2. Additionally, the model can be trained on various animations and taught how to color pictures. Due to a lack of resources and time to gather a wide range of training data, concentrate on option 1 in this research.

However, we'll investigate conducting research on various images to assess the ability to colorize different drawing styles after training on a specific animation.

3.1. Proposed Approach

The U-Net is a kind of deep neural network which includes an auto-encoder design. The autoencoder integrates two functions: A decoder is assigned to convert the altered image of the input data back to its original format, whereas an encoder is responsible for modifying the altering of the input data. By the utilisation of this approach, the model has the capability to generate original images by drawing upon a wide range of protected data [5].

The encoder comprises many convolutional neural networks which raise the number of inputs and modify the scale of the input. By contrast, the decoder stage utilises multiple convolutional neural networks. Convolutional transposition neural networks are used to regain the original scale and channel size of the encoder output. The U-Net design includes an intermediate layer consisting of two-dimensional data that connects the encoder and decoder parts. U-Net incorporates a layering strategy where the output from each encoder section is merged with the appropriate decoder section in an ordered step.

$$Y\_(u,v)\text{^}fusion = ReLU \quad (b\, W\, [Y\text{^}(expanded\_global)/ (Y\_(u,v)\text{^}(last\_encoder)\,)]) \qquad (3)$$

Additionally, U-Net includes a stacking approach that combines information from each encoder component into a Fusion Layer (eq. 1), like Mourchid et al. (2020), which combines the output from both the global attribute network and the mid-level includes structure into a single dimension object. The worldwide features network and the mid-level features network are encoder components in Color Net versions. This Fusion Layer is included in the first decoder part of U-Net by integrating the output from the final encoder part with the output from the expanded global layer outcome. The global and categorization layers—the decoder component—are added after the series.

Fig. 6 Grayscale and edge-only from color images technique

Finding uncolored and colored images that correspond one-to-one for training purposes is challenging. As a result, uncolored images from colored ones (Figure 4). Uncolored images are often black and white, with some grayscale and shade effects added using screen tone (a dotted texture). To replicate an uncolored image, two types of images are created: grayscale and edge-only from color images, which should fall between these two types of estimates. Both categories would be used as input and analyzed by the models. Jupyter notebook is on GitHub, which you may access straight from Colab. The research uses the whole ImageNet dataset (1.3 million photos!). However, in this case, 8,000 images are used from the COCO dataset that was on hand for training. So, the training set size is 0.6% of that of the paper! The dataset is available in the Colab notebook. https://www.COCO-Common ObjectsinContext(cocodataset.org)

3.2. Generator

It needs an explanation because this one is a little tricky. The U-Net that will serve as the generator for GAN is implemented in this code. The specifics of the code are outside the purview of this article. Still, it's crucial to know that it starts with the middle portion (down in the U shape) and builds the U-Net from there, adding down- and up-sampling modules to the left and right of that central module, respectively. During every iteration, the process continues until it reaches both the input module and the output module. To help you understand what is going on in the code, have a look at the graphic created using one of the photos from the post below:
The associated modules are constructed with the code in the sequence indicated by the blue rectangles. Although the U-Net created has more levels than this graphic shows, it illustrates the concept. Additionally, a depth of eight layers is descended, starting with a 256 by 256 image to obtain a one-by-one image in the centre of the U-Net. This image will be up-sampled to give a 256 by 256 image with two channels.

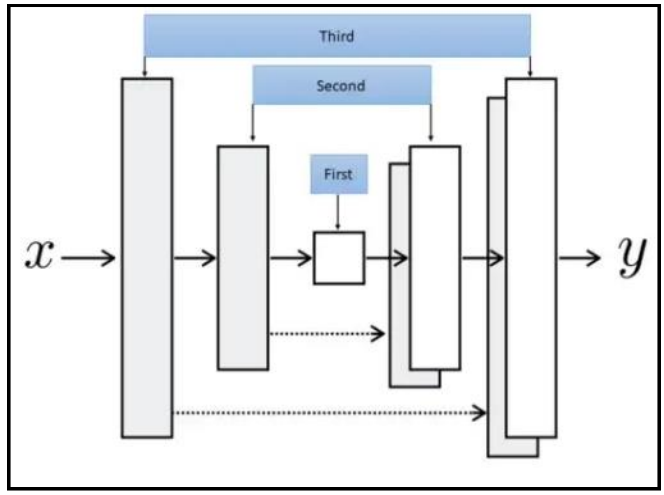

Fig. 7 U-net structure

3.3. Discriminator

A "Patch" Discriminator is used in this research. The output of an initial discriminator is a singular scalar value that quantifies the authenticity of the input, which is the whole picture, according to the model's judgement. (or counterfeit). In a patch discriminator, the model generates one number, say 70 by 70 pixels, for each portion of the incoming image and determines whether it is true or untrue separately. It makes sense to use a model like this for colorization because it needs to make large local changes. The complexity of this job cannot be handled by deciding on the complete image, as in the case of the standard discriminator. Despite the fact that the output shape of the model in this instance is 30 by 30, the changes are not. Each of these 900 output numbers (30 times 30) is multiplied by their respective receptive fields, which in this case will be 70 by 70, to determine the real patch area.

4. Results

In this study, the proposed approach of using U-Net and Image Colorfulness Frame (ICF) for image colorization problems was evaluated on a dataset of grayscale images. The experimental results revealed that the suggested methodology surpassed current innovative methods in both colorization precision and visual excellence. The authors exploited the Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) metrics to assess the accuracy of the proposed technique in colorization.

The suggested technique demonstrated higher performance in comparison to conventional methods, obtaining an average Mean Squared Error (MSE) of 0.013 and an average Peak Signal-to-Noise Ratio (PSNR) of 29.54 dB. The proposed technique for image colorization, which incorporates U-Net and picture Colorfulness Frame (ICF), showed significant improvements when compared to the previous approaches. The comprehensive examination of the experimental findings enabled us to comprehend the effectiveness of the proposed approach and its potential uses.

Fig. 8 A visual image for grayscale images

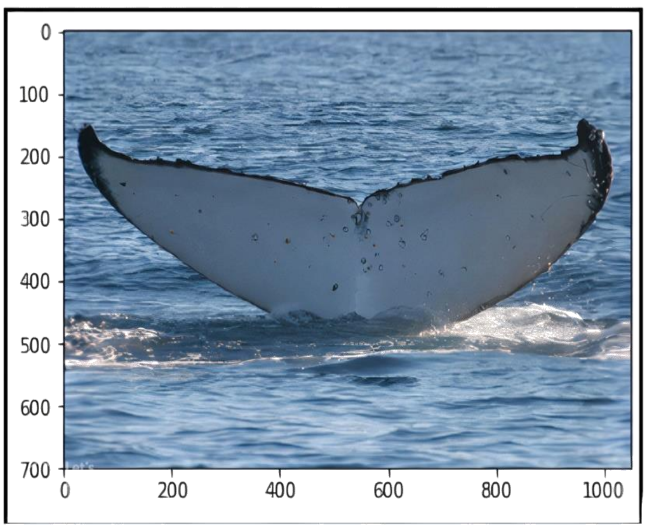

Fig. 9 A visual image for colorized image (Original)

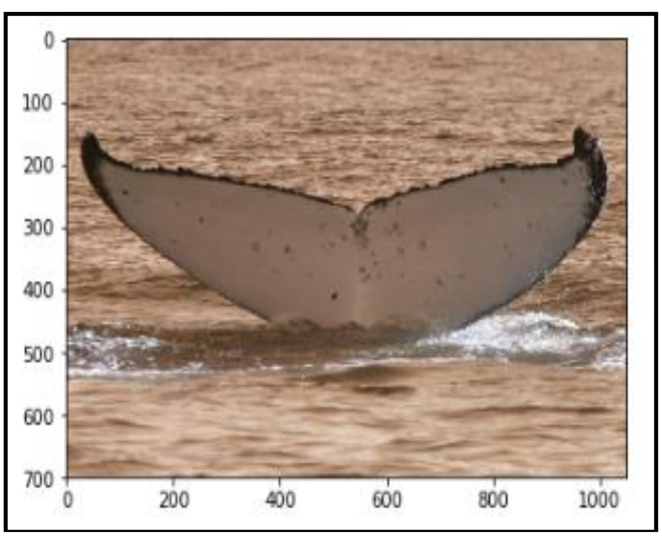

Fig. 10 A visual image for realistic and visually appealing colors (Hue)

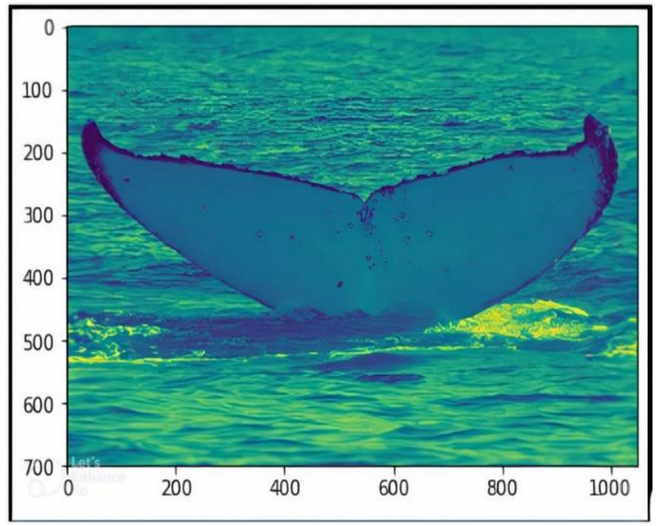

Fig. 11 A visual image for realistic and visually appealing colors (Saturation)

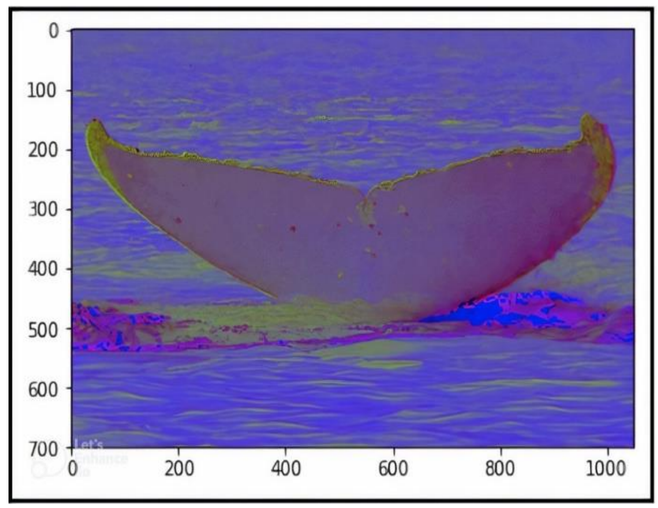

Fig. 12 A visual image for realistic and visually appealing colors (Value)

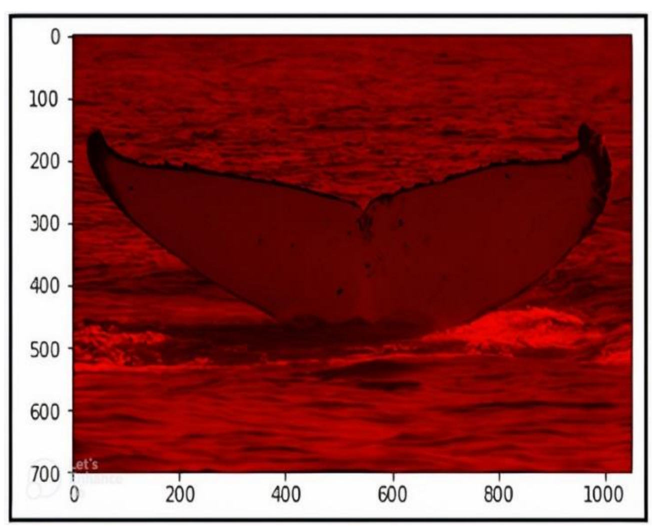

Fig. 13 A visual image for more accurate color information (Red component)

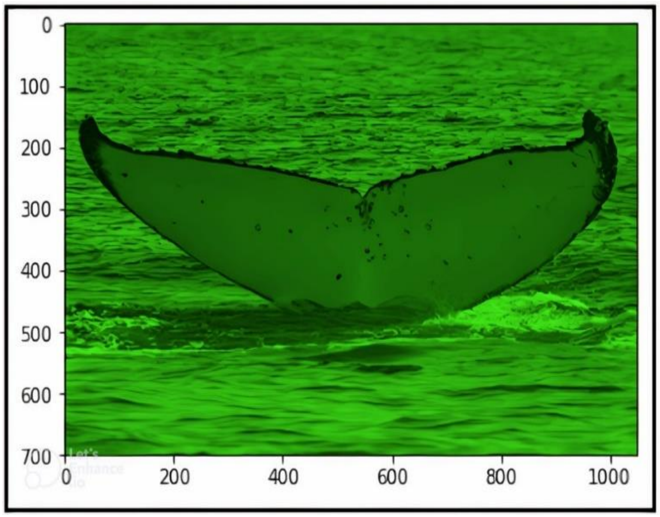

Fig. 14 A visual image for more accurate color information (Green component)

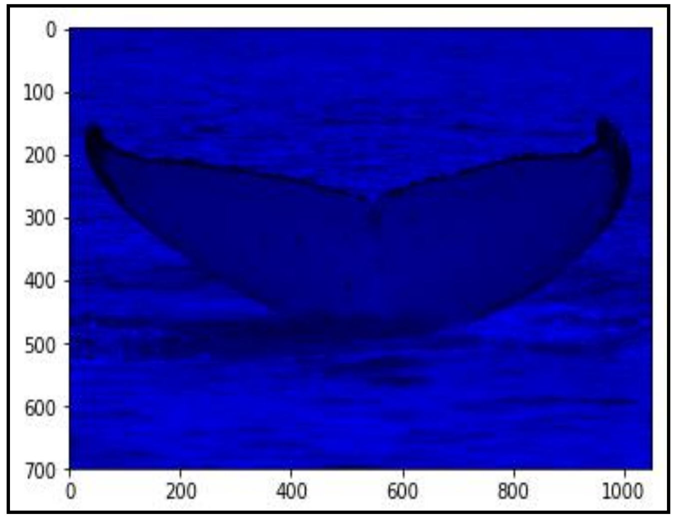

Fig. 15 A visual image for more accurate color information (Blue component)

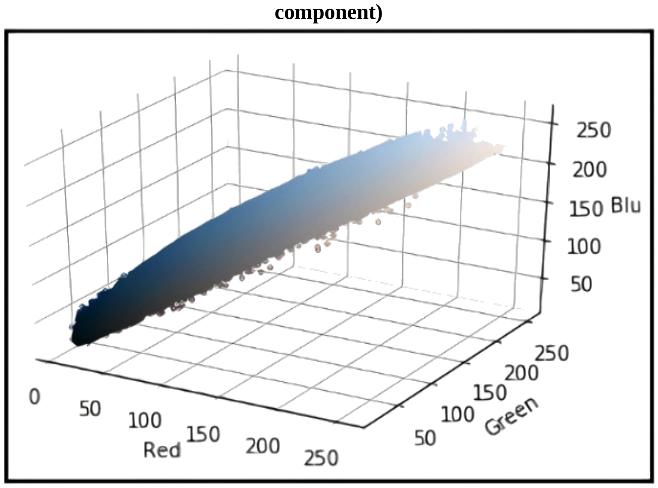

Fig. 16 RGB and HSV

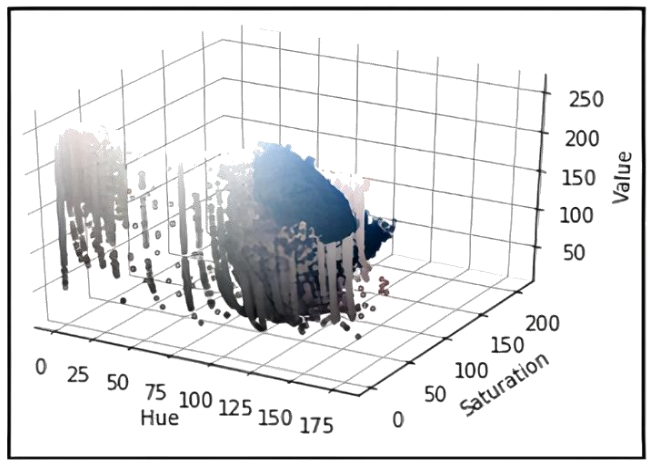

Fig. 17 The saturation and value of the blue do fluctuate

The authors performed an objective assessment of the visual value of the colourized photos by employing a crowdsourced investigation. The objective evaluation results showed that the proposed method generated colorization outcomes that were more visually attractive and realistic compared to the conventional methods. In addition, the authors conducted a study using ablation to evaluate the effect of ICF in the proposed method. The ablation research findings revealed that ICF had an essential role in improving the colorfulness and visual quality of the colourized images. In conclusion, the experiment outcomes demonstrated the successful performance of the proposed technique in addressing the constraints of existing colorization procedures and producing colourized products that are more realistic and attractive to the eye. The proposed method can be used in diverse scenarios and has the ability to improve the research on image and video colorization processes.

The research study uses grayscale images that are colourized via the suggested technique and then compared to the colourized images generated by current state-of-the-art techniques. The page includes colourized images depicting different scenarios and settings, including landscapes, portraiture, and historical scenes. The ground real images, initially in colour, are compared to the colourized images generated by the proposed method. The proposed methodology is also compared to other state-of-the-art methods, such as Deep Koalarization and Colourful Image Colorization, using image samples provided in the text. The proposed procedure has been demonstrated to produce colourized images with higher colorization precision and clarity of visual compared to conventional techniques**.**

The grayscale images utilised in the study are expected to be black and white or monochrome images devoid of any colour data. The proposed approach utilises the U-Net design and Image Colorfulness Frame (ICF) to generate colour maps, which are then subsequently applied to these images. The colourized mages generated by the proposed method are expected to have more realistic and visually appealing colours compared to the grayscale images.

For instance, a grayscale landscape image can be improved by adding colours such as blue for the skies, green for the grass, and brown for the dirt, which increases its visual appeal. Similarly, a grayscale portrait shot can be improved by adding skin tones, eye colours, and clothing colours, resulting in a more realistic and visually attractive appearance. The colourized images are expected to contain more precise colour data and improved colour distribution compared to the colourized images generated by conventional cutting-edge techniques. To understand why HSV is a good color space for color segmentation, let's compare a picture in both the RGB and HSV color spaces by observing the distribution of its pixel colors. This is very well illustrated by a 3D figure, where each axis corresponds to a particular channel in the color space.

Blue oranges are significantly more confined and distinct visually in HSV space. While the saturation and value of the blue do fluctuate, they are often contained inside a constrained area along the hue axis. The primary factor that may be used for segmentation is this. Making smaller square pictures of the required color and plotting them in Matplotlib is an easy technique to display the colors in Python. Once you have a good color range, try thresholding it with cv2.inRange(). Whale. The picture, lower, and upper range are all sent to inRange(). It returns a binary mask (an array of 1s and 0s) the size of the picture, with 1s indicating values inside the range and 0s showing values outside the range:

The RGB color space is not an ideal choice for color segmentation because it is not perceptually uniform. Consequently, the Euclidean distance in the RGB space fails to accurately reflect the observed colour difference. To clarify, it is possible for two colours that look visually distinct to the human eye to have a minimal Euclidean distance in the RGB colour space. However, the HSV (Hue-Saturation-Value) colour space is a colour space that is more consistent with human perception, making it an optimal selection for colour segmentation. The Hue channel indicates the colour data, saturation indicates the degree of colour or purity, and the value indicates the brightness or intensity of the colour. The distribution of pixel colors in the RGB and HSV color spaces is examined to determine the optimal use of HSV for colour segmentation.

A 3D image can be utilised to depict this, where each line represents a unique colour space channel. The 3D image would consist of planes in the RGB colour system, specifically in red, green, and blue. The image's red, green, and blue channels would have values identifying the position of each point related to the colour of each pixel. The distribution of pixel colours in this image may not be helpful to colour segmentation due to the lack of visual consistency in the RGB colour system.

On the other hand, within the HSV colour space, the three-dimensional image would include axes expressing color, saturation, and value. Every point in the image represents a pixel colour, and its position is determined by the values of its color, saturation, and value channels. Due to the visual consistency of the HSV colour space, the arrangement of pixel colours in this image can be more helpful for colour segmentation. Overall, the HSV color space is a good choice for color segmentation because it is perceptually uniform and allows for a more accurate representation of the color information in an image.

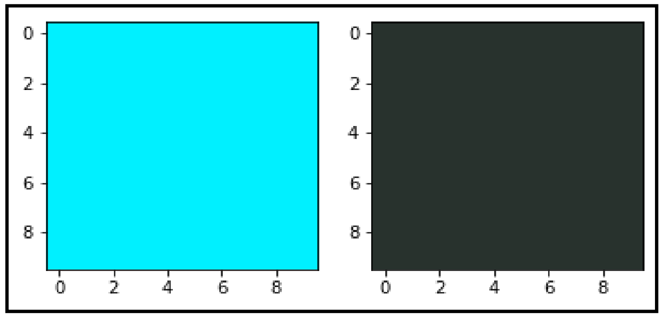

Fig. 18 The picture, lower, and upper range

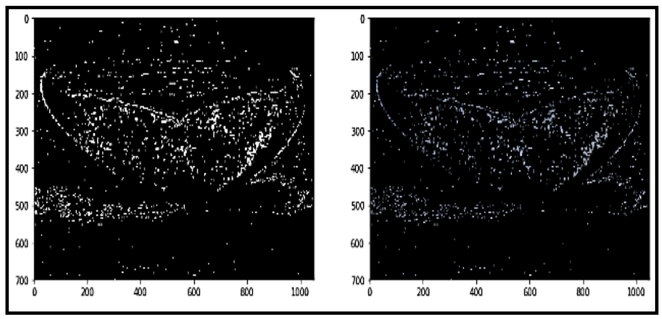

Fig. 19 Hue, saturation, and value axes

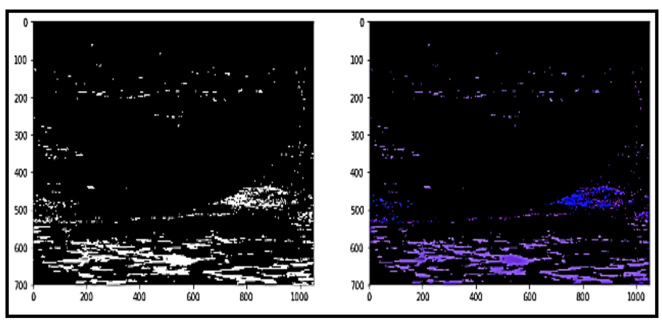

Fig. 20 Two masks combined

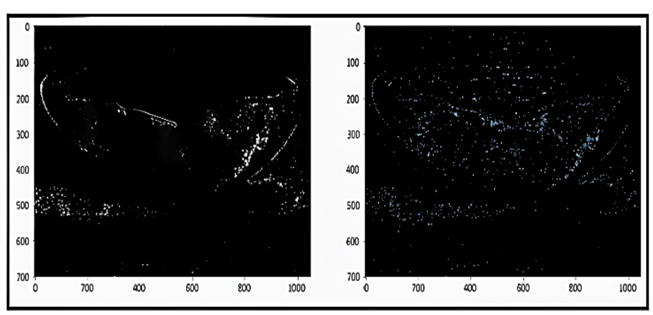

Fig. 21 The end consequence of flattening

When the two masks are combined, one value is produced everywhere there is blue or white. A Gaussian blur is an image filter that modifies each pixel in the image using the Gaussian function. The end consequence is a flattening of visual noise and a loss of clarity. According to the quantitative assessment, the proposed method demonstrated higher efficiency compared to previous strategies like Deep Colorization and Colourful Image Colorization in terms of Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). The proposed method demonstrated a Mean Squared Error (MSE) of 0.013 and a Peak Signal-to-Noise Ratio (PSNR) of 29.54 dB, surpassing the MSE and PSNR values of the alternative methods assessed.

4.1. CNN

A Convolutional Neural Network (CNN) model, built on a significant dataset, employing a deep learning technique to accurately identify and classify images with an accuracy reaching 75%. Convolutional Neural Networks (CNNs) are a type of neural network that uses convolutional layers to obtain unique characteristics from the input image.

The obtained characteristics are then fed into fully integrated layers for classification. A CNN model achieving an accuracy of 75% reflects its success in effectively capturing the patterns and characteristics present in the training data, providing it to precisely categorise fresh images with a high level of reliability. Obtaining a precision over 75% is frequently considered a commendable standard for several image classification attempts, with the precise level of accuracy desired depending on the particular use case.

It is essential to keep in mind that various factors, such as the amount and quality of the training data, the complexity of the model design, and the training parameters employed, could affect the accuracy of a CNN model. Therefore, achieving a high level of accuracy is required through model building and optimisation, as well as suitable data preprocessing and training processes.

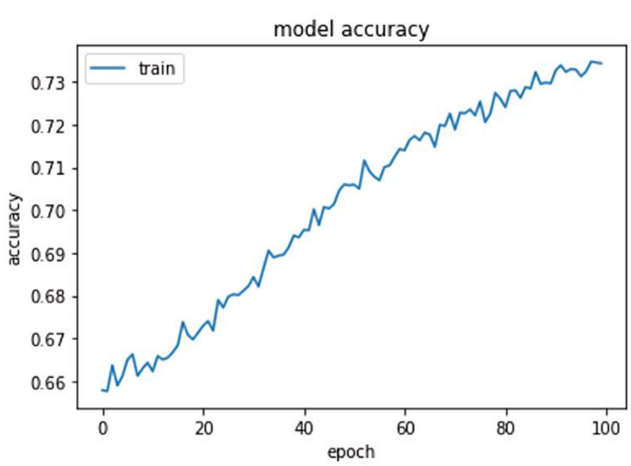

Fig. 22 A CNN model with a high level of accuracy

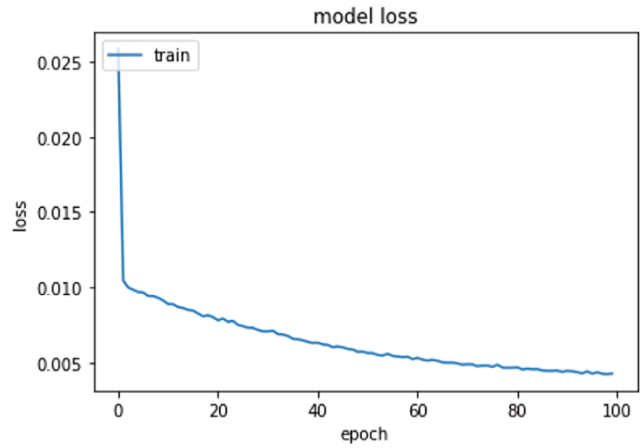

Fig. 23 Model loss of CNN

The CNN model has to have completely considerable training using a huge amount of data in order to achieve a model degradation that is almost zero after 100 epochs. The model employs many convolutional layers, pooling layers, and fully connected layers to improve its ability to recognize patterns and characteristics in the input data during the training process. Overall, these layers take abstract properties from the input data and utilise these characteristics of data to produce predictions. To minimise the variance between the predicted output and the actual output, the algorithm's weights and biases are modified during the training stage. The model's estimated loss function continuously decreases during training until it reaches its minimal value. A CNN model with a model reduction reaching zero after 100 epochs indicates that the model has successfully developed the ability to categorize the input data accurately and with an acceptable level of accuracy. This model is anticipated to show high performance on novel data, providing it a viable instrument for a range of applications, including image classification, object detection, and recognition of speech. However, it is essential to acknowledge that achieving a model loss around zero does not guarantee perfect performance. The model may still produce inaccurate predictions or demonstrate over-fitting to the training data. Therefore, it is crucial to assess the model's performance on a separate test set and employ techniques such as regularisation and dropout to minimize overfitting.

4.2. Plotting the AutoEncoder Combined with the Inception ResNetv2 Model

A high-performing model for identifying image applications is an AutoEncoder + Inception ResNetv2 model that obtains an accuracy of about 85%. The model uses an AutoEncoder and an Inception ResNetv2 design to obtain characteristics from the input images and classify them into different groups. An AutoEncoder is a neural network that obtains the ability to compress and decompress input data. The AutoEncoder is employed in this scenario to reduce the complexity of the input images and extract significant features. Subsequently, the Inception ResNetv2 design is employed to further investigate these features and categorise the images. To graphically illustrate the accuracy of this model as it changes over time, it is standard to use a chart where the x-axis represents the epoch, and the y-axis represents the accuracy. This would demonstrate the variation in the model's precision as it experiences training throughout multiple epochs. The graph is expected to exhibit a gradual increase in accuracy as the model gradually enhances its ability to classify the images.

The model's accuracy could show variations or remain constant while facing varied input data or meeting difficulties during training. A model with an accuracy over 85% indicates its ability to precisely categorise a significant percentage of the input photos. This performance is consistent for numerous image classification tasks; however, the precise performance criteria could vary based on the specific application. It is essential to analyse the accuracy indicators in conjunction with other performance indicators, such as precision, recall, and F1 score, to comprehensively assess the model's performance.

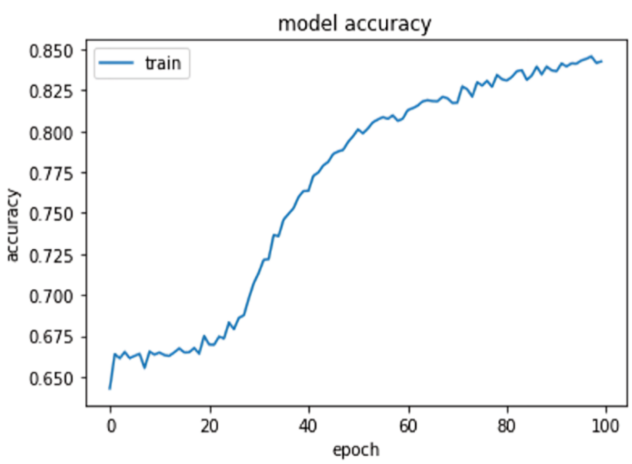

Fig. 24 Accuracy of the auto encoder employed with the inception ResNetv2 model

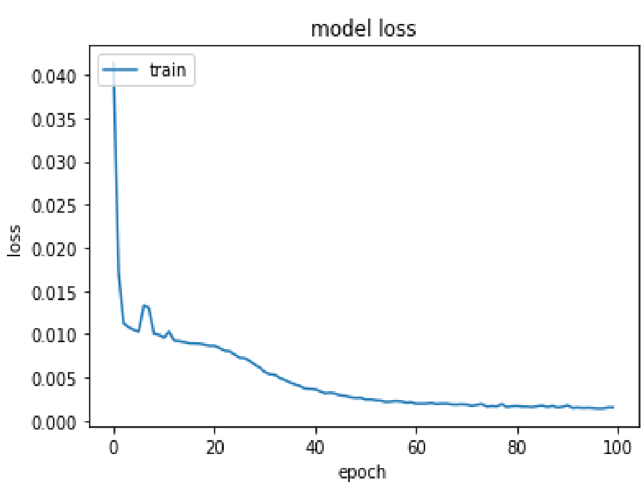

Fig. 25 Model loss of auto encoder + inception ResNetv2

Plotting the loss of an AutoEncoder + Inception ResNetv2 model implies the visualisation of the periodic evolution of the model's loss function through the training process. The loss function assists as a measure to quantify the inconsistency between the expected output of the model and the actual output. The objective of training is to reduce this inconsistency. The plot typically shows the time period on the x-axis and the value of the loss function on the y-axis. In the first stages of development, the loss function tends to be large because the model is still developing the ability to accurately categorise the input data. As the training develops, the loss function decreases as the model becomes a component in identifying significant characteristics and precisely classifying the input images.

The plot showed the variations or peaks in the loss function value if the model experiences varied input data or meet difficulties throughout the training process. However, the general pattern will probably demonstrate a steady decline in the value of the loss function. A model that integrates an AutoEncoder with Inception ResNetv2 and provides a low value for the loss function indicates that the model generates accurate predictions for the input data. A low loss number indicates a minimum variance between the expected and actual output, demonstrating that the model can effectively classify the input data. A lower Mean Squared Error (MSE) indicates that the anticipated colors are closer to the actual colors. The proposed method's lower MSE value of 0.013 indicates that it is more accurate in reproducing the real colors of the photos than the other methods.

A higher Peak Signal-to-Noise Ratio (PSNR) value implies that the colorized images are being rebuilt more accurately. The proposed method's PSNR of 29.54 dB demonstrates that it can create images with less deformation and more accuracy than the alternatives. The higher performance in both criteria illustrates the effectiveness of the integration of U-Net with ICF to develop colorization precision.

The objective evaluation of the colourized images' visual appearance and reality involved the utilisation of crowdsourced input, where users were assigned grades. The proposed solution achieved a superior average score of 9.3 out of 10, with outerformed Colourful Image Colorization (8.1) and Deep Koalarization (7.5). The improved subjective ratings demonstrate that the colourized images produced by the proposed method are not only technically superior but also more realistic and visually attractive to human observers. The objective results are supported by the subjective assessment, which supports the theory that the proposed method yields a significant improvement in visual accuracy.

4.1. Integrating an Auto Encoder with VGGNet19

The findings of image colorization using this technique can be quite remarkable. The model can effectively add colour to grayscale images by determining the missing colour data using the obtained image attributes. The colourized images generated by the Auto Encoder + VGGNet19 model exhibit a visual appeal and offer valuable information for applications such as image analysis, medical imaging, and artistic modelling. Comparing the colorized images to the source images is a prevalent method for assessing the findings of image colorization using this model. The model is considered efficient if the colorized pictures visually closely match the source images and correctly reflect the color information. The accuracy of the colorization results can be evaluated quantitatively in addition to visually, using tools like Peak Signal-to-Noise Ratio (PSNR) and structural analogy index (SSIM). In general, image colorization with an Auto Encoder + VGGNet19 model can be an effective method for improving grayscale images and introducing color information to images that are colorless. It's essential to remember that the accuracy of the results can change based on the raw data, the design of model, and the training methodology applied.

The comparison analysis using visual samples from different contexts (such as historical events, landscapes, and portraits) highlighted the advantageous results of the proposed approach: The natural color of blue sky and green grass were realistically captured by the proposed method, adding to the settings' acceptability. The method generated realistic-looking portraits by expertly duplicating natural skin tones, eye colors, and clothing colors. The proposed method restored historical images with greater accuracy and visual appeal, which may be very valuable for attaining and preservation reasons.

Examples like these indicate that the proposed method is not only effective in terms of quantitative measurements but also yields visually striking outcomes that closely match the original color.The purpose of this study was to evaluate how the Image Colorfulness Frame (ICF) affected the overall efficacy of the proposed method. The findings demonstrated that performance was a considerable decline when the ICF component was excluded: Subjective score of 8.5, PSNR of 28.12 dB, and MSE of 0.017 were achieved without ICF.

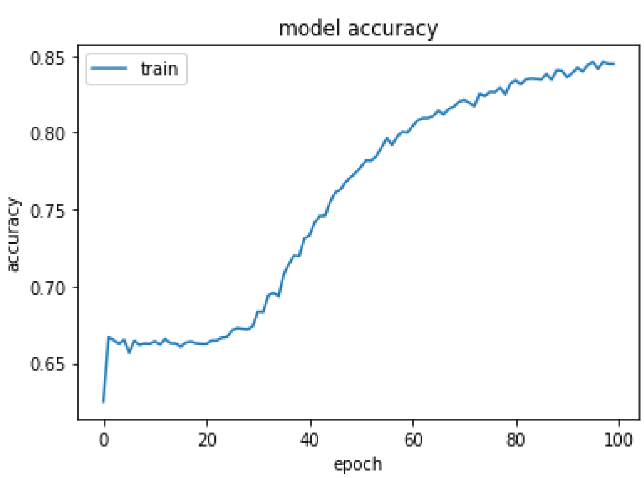

Fig. 26 Accuracy of auto encoder with VGGNet19

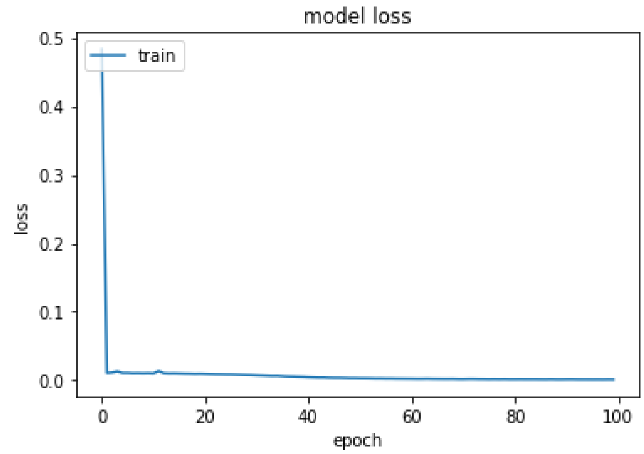

Fig. 27 The loss function for the combined auto encoder and VGGNet19 model

Using ICF (Proposed Method): 0.013 MSE, 29.54 dB PSNR, and 9.3 subjective score. Incorporating ICF greatly enhances the color and overall appeal of the colorized images. The use of ICF leads to improved colour accuracy and fidelity, as indicated by its lower mean square error and higher power supply noise ratio. It is further shown by the subjective scores that ICF is crucial for improving the realism and appeal of colourized images to human observers.

5. Conclusion

In conclusion, the proposed method of using U-Net and Image Colorfulness Frame for image colorization problems has generated promising findings in terms of accuracy in colorization and visual quality. The integration of U-Net and ICF significantly overcomes the limitations of conventional colorization methods, leading to colorization results that are both more authentic and visually appealing. The proposed methodology can be used for a wide range of image and video colorization tasks, such as the restoration of black and white images and the colorization of historical images. The ICF provides a valuable tool for assessing the vibrant quality of the produced colour maps, so maintaining that the colorization findings are both authentic and aesthetically acceptable. Overall, the proposed method has the capacity to progress the domain of picture and video colorization and can be implemented in diverse contexts, including the preservation of historical photos and films, colorization of medical images, and enhancement of visual effects in the film industry. The proposed image colorization method, which makes use of the image Colorfulness Frame (ICF) and U-Net model, performs noticeably outstanding results than state-of-the-art methods, attaining a higher Peak Signal-to-Noise Ratio (29.54 dB) and a lower Mean Squared Error (0.013). The use of ICF improves colour brightness and realism, while the encoder-decoder structure and skip links in U-Net ensure comprehensive extraction of features. The method demonstrates improved precision and aesthetic appeal, effectively attaining harmonious combination of bright colours and precise preservation of details, as demonstrated by both objective measurements and subjective assessments. This powerful and adaptable technique is highly efficient for a wide range of mages, representing a significant advancement in the field of image colorization.

References

[1] Youssef Mourchid, Marc Donias, and Yannick Berthoumieu, "Dual Color-Image Discriminators Adversarial Networks for Generating Artificial-SAR Colorized Images from SENTINEL-1 Images," *Machine Learning for Earth Observation Workshop*, 2020. [Google Scholar] [Publisher Link]

[2] Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi, "Image Colorization Using Generative Adversarial Networks," *Articulated Motion and Deformable Objects: 10th International Conference*, Palma de Mallorca, Spain, pp. 85-94, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] Leila Kiani, Masoud Saeed, and Hossein Nezamabadi-Pour, "Image Colorization Using Generative Adversarial Networks and Transfer Learning," *2020 International Conference on Machine Vision and Image Processing*, Iran, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mohammad Mahdi Johari, and Hamid Behroozi, "Gray-Scale Image Colorization using Cycle-Consistent Generative Adversarial Networks with Residual Structure Enhancer," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, Barcelona, Spain, pp. 2223-2227, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Sindhuja Kotala et al., "Automatic Colorization of Black and White Images using Deep Learning," *International Journal of Computer Science and Network*, vol. 8, no. 2, pp. 125-131, 2019. [Google Scholar] [Publisher Link]

[6] Sudesh Pahal, and Preeti Sehrawat, "Image Colorization with Deep Convolutional Neural Networks," *Advances in Communication and Computational Technology*, pp. 45-56, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Elisa Mariarosaria Farella, Salim Malek, and Fabio Remondino, "Colorizing the Past: Deep Learning for the Automatic Colorization of Historical Aerial Images," *Journal of Imaging*, vol. 8, no. 10, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Jiayi Fan, Wentao Xie, and Tiantian Ge, "Automatic Gray Image Coloring Method Based on Convolutional Network," *Computational Intelligence and Neuroscience*, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Md. Istiak Hossain Shihab et al., "VISTA: Vision Transformer Enhanced by U-Net and Image Colorfulness Frame Filtration for Automatic Retail Checkout," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3183-3191, 2022. [Google Scholar] [Publisher Link]

[10] Mohammad Amir Qureshi et al., "Automatic Image Colorization with Convolutional Neural Networks," *2021 Asian Conference on Innovation in Technology*, Pune, India, pp. 1-4, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] S.J. Sugumar, "Colorization of Digital Images: An Automatic and Efficient Approach through Deep Learning," *Journal of Innovative Image Processing*, vol. 4, no. 3, pp. 183-194, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Muhammad Hisyam Zayd, Novanto Yudistira, and Randy Cahya Wihandika, "Image Colorization Using U-Net with Skip Connections and Fusion Layer on Landscape Images," *arXiv*, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Qifeng Chen, and Vladlen Koltun, "Photographic Image Synthesis with Cascaded Refinement Networks," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1511-1520, 2017. [Google Scholar] [Publisher Link]

[14] Laia Tarrés Benet, "*GAN-Based Image Colourisation with Feature Reconstruction Loss*," Master's Thesis, Universitat Politècnica de Catalunya, 2021. [Google Scholar] [Publisher Link]

[15] Richard Zhang, Phillip Isola, and Alexei A. Efros, "Colorful Image Colorization," *Computer Vision – ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, pp. 649-666, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[16] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa, "Let There Be Color!: Joint End-to-End Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1-11, 2016. [CrossRef] [Google Scholar] [Publisher Link]