An approach to pseudocoloring of grey scale image using deep learning technique

Kshitija Srivastava[1], Saksham Gogia[1], and Rohith G[2]

[1]Student, School of Electronics Engineering, Vellore Institute of Technology, Chennai 600 127, India
[2]Assistant Professor Senior Grade[1], School of Electronics Engineering, Vellore Institute of Technology, Chennai 600 127, India

Abstract. Image pseudo colorization is the process of adding RGB colours to grayscale images to make them more appealing. Deep learning technology has made progress in the field of automatic colouring. In general, we divide automatic colouring methods into three groups based on where the colour information comes from: colouring based on what you already know and on reference pictures. The colouring method can meet the needs of most users, but there are some drawbacks. For example, users can't colour different reference graphs for the different things in a picture. In order to solve this problem by recognising several objects and background regions in a picture and combine the final colouring results, the proposed method uses the deep learning approach that regional mixed colours be used more and the method be mastered by using deep learning. Qualitative results (visual perception) validate the effectiveness of pseudocolorisation which split into foreground colour based on a reference picture and background colour based on prior knowledge. Quantitative results such as Structural Similarity (SSIM), Peak Signal to Noise Ratio (PSNR), Image Matching Error and Entropy validates the effectiveness of strong edge strength, visually appealing quality and retention of maximum information without disturbing quality of image.

1. Introduction

Pseudocolorizing pictures is difficult because of the wide range of imaging situations that must be addressed by a single algorithm. Although the scene semantics may be useful in many cases—grass is often green, clouds are typically white, and the sky is typically blue—two of the three picture dimensions are absent, making the problem significantly ill-posed. In contrast, many man-made and natural items, such as clothing, vehicles, flowers, etc., seldom ever have such semantic priors. Furthermore, the colorization problem inherits the standard difficulties of picture improvement, including shifts in lighting, shifts in perspective, and occlusions. After the creation of an animation's line art draught image, the colouring process is a crucial but labour-intensive next step. Modern cartoonists, in an effort to cut down on production time, will use specialised business equipment and software to speed up the colouring process. The research topic of single-image pseudocolorization is of crucial importance uses in the actual world. The unique advantages of deep learning techniques Because of this, deep convolutional neural networks have expanded rapidly ways of adding colour to images. Founded on promising new developments, Networks are the focus of a number of suggested methodologies, and these frameworks, instructional strategies, pedagogical modalities, etc. Here, we see that the quality of image colorization has increased. At the expense of expanding the network in recent years complexity. However, the use of cutting-edge techniques. Currently, our ability to respond to crucial real-world circumstances is constrained by insufficient complicated networks, low-quality measurements, and an inability to degradations in the actual world. Over the past few years, with to rapid advancements in computing technology, the concept of artificial intelligence has increasingly entered the public consciousness. Using automated colorization for images has the potential to not only pique people's interest in mundane tasks like producing posters, but also to boost productivity in the same endeavours. As a result, scientists are also interested in grey picture colouring technologies. Using colour images as a reference, the authors of papers [1]-[3] propose a new automatic colouring method for line art images that can be applied to areas of a similar colour. Researchers use deep convolutional neural networks (DCNNs) and propose many methods for automatic colouring of line draught images, inspired by the successful application of generative models in image synthesis tasks in recent years; however, the colouring results of these methods are not controllable and are often accompanied by colour artefacts. The digital media industry, including film and television advertising and online games, has flourished in recent years thanks to advances in Internet technology, and in order to draw in an audience, these mediums typically require eye-catching posters [4]-[6]. The quality of a work can be judged not only by its layout, but also by the way its colours work together. Colorization is a crucial step in any media production, from still images to motion pictures. However, this is no simple feat, as the selection and placement of colours is both time-consuming and a litmus test of the artist's technical proficiency and aesthetic maturity. Furthermore, if the colour of one of the objects in a good work is not satisfactory and recoloring is required, the current method is direct grey recolor, which is a massive undertaking. Therefore, multiarea image colouring is an important area of study for both academics and professionals. Recently, several different picture colorization models have been created, with state-of-the-art performance being reported on modern datasets, all thanks to the advancements in deep learning. Different deep-learning models have been employed to solve the colorization issue, from early brute-force networks to the more recently meticulously built Generative Adversarial Networks (GAN). There are numerous key distinctions among these colorization networks. An example of a simple and effective neural network is the convolutional network that's radically distinct from all others in deep learning models. Models for deep learning in general include a global. The structural mode of a neural network is created by CNN. It include the generator sub network and the receiver subnetwork. The discriminator-specific sub network, and use the generator for extract features from images and create fabricated versions. Discriminators can be used to tell the difference between authentic and counterfeit images, calculating the likelihood that the image is greater potential for authenticity or image generation. That's because of this, where the procedure, including the discriminator model as well as the generator model training the model all the time [7]. As the number of repetitions grows, the forged images produced by the generator model. The discriminator's ability to do so also increases. Capability of Discerning between Real and Fake Pictures increases in intensity and tends to converge at the end. In light of this, the CNN model is commonly employed in the field of image processing and is arguably the most popular framework for domain of picture colouring [8]–[11]. The process of colorization aims to assign RGB values to a grayscale image, which was likely taken before widespread access to colour cameras and subsequent technical improvements. As such, this method is more akin to picture enhancement than it is to image repair. Image colorization is also useful for restoring colour to photos that were converted to grayscale or the Y-Channel of YUV color-space so that more room or bandwidth might be made available for other data. As a result, a simple formula for this situation looks like this:

$$I_g = \Theta(I_{rgb}) \quad (1)$$

where $\Theta(.)$ is a function that converts the RGB image I_{rgb} to a grayscale image Ig.

$$I_g = 0.2989 \times I_r + 0.5870 \times I_g + 0.1140 \times I_b \quad (2)$$

As opposed to the three channels required in RGB, colorization techniques typically try to restore colour in YUV space, where the model only has to forecast two channels, U and V.

The significant contributions of the proposed work are

- We suggest a modified CaffeNet model to increase the usage of regional mixed colour and master the approach by applying it to the context of colour picture segmentation and image fusion technology. It may be broken down into two categories: foreground colour (determined by the reference image) and background colour (determined by the user's past knowledge).We employ the 1.2 million high-resolution training pictures from the ImageNet ILSVRC2012 [12] dataset.
- To the best of author’s knowledge, the proposed model is first of its kind and hence comparison is not performed. Moreover, the image matching error metric is proposed.

2. Related Works

Chen et al. [13] turned to conditional generative adversarial networks (cGANs), which allow for automatic colouring with little to no human intervention. The approach works well for understanding the connection between grayscale and colour images, but not line images. In order to continually update the learning model parameters and increase the model's performance for unlabelled region classification labels, an active learning framework learns domain classification labels in small data sets and assists users in selecting the data to be classified in unlabelled sets. In contrast to prior approaches of picking marked data based on uncertainty, Zeng et al. [14] introduced an adaptive active learning approach that coupled information density calculation with least uncertainty calculation to choose marked instances. Users were requested by An et al. [15] to create a colour curve for graffiti and determine the gradient range of the curve in order to limit the spread of vandalism. To get the final picture, the approach solves Poisson's equation under the restrictions of a set of diffusion curves. However, all of these approaches rely heavily on user input to get the desired hue. Researchers presented a colouring approach based on reference images to cut down on manual labour and allow for the realisation of the required colour style. Using a graph representation of the connections between the various parts of a line drawing image, JWA et al. [16] found a quadratic programming solution to the matching issue. However, it is typically challenging to effectively segment a complicated line-drawing image, leading to many blocks being created for the same semantic region. In order to circumvent the need for precise image segmentation, researchers have recently suggested a deep learning image colouring method based on a line-drawing map guided by reference images. In order to colourize black-and-white photos. The colouring challenge was introduced by Farid et al. [17], and it requires deriving three-dimensional information, such as the RGB channel, from the one-dimensional information (intensity or brightness) of a grayscale image. The correspondence between the dimensions of the data and the dimensions of the map is not special. Because of the inherent ambiguity of colorization, it is necessary to include sufficient contextual context. Therefore, the speed of the process may be improved, and the quality of the static picture can be maintained using a colouring technique based on image brightness-weighted colour mixing and rapid feature space distance computation. For comics, Berger et al. [18] presented a new automated colouring technique.

There are both global and local aspects to an image. Kotecha et al. [19] used a deep network to learn the fine-grained properties of colour pictures in order to train an automated system for colorization of black-and-white photographs. This model was then used to predict the colour information of each pixel in the black-and-white images. The image's global characteristics consist of its overall shape and contours, whereas its local features are its individual parts. The network, which is based on a convolutional neural network, is divided into two sub networks—a local feature extraction network and a global feature extraction network—to analyse cartoon images of arbitrary resolutions. Recent years have seen a shift in the mainstream of cartoon colouring toward a synthesis of two approaches: the realisation of a basic colouring algorithm and a deep learning model. Image translation is one of the many tasks that have been successfully completed using deep learning-based technologies. Many studies have attempted to employ human cognition in the quest to automate the colouring of photographs. In a recent work, A. Sakur et al. [20] noted the dearth of unsupervised learning-based research on image processing. They looked at the CNN network and came up with DCGAN, a deep volume network, to help with both supervised and unsupervised learning.

Numerous studies have been conducted on the topic of picture region segmentation, as the precision with which the image is sliced is a crucial component in determining the final image's colour accuracy and quality. To improve the visual impression, Oladi et al. [21] took into account not only neighbouring pixels with comparable intensity but also faraway pixels with the same texture. Results improved when pixels were coloured near edges based on texture similarity and pixels in smooth regions based on intensity similarity, according to preliminary testing. They have built a set of interface tools that enable users to tag, colour, and change target pictures; thus, the technique may even be used to colour comics. Two U-NET-based colouring approaches were implemented by Qiao et al. [22], with the following improvements over prior work: Moreover, the network will also supply users with a data-driven colour palette, proposing the optimal hue of the grey map at a specific place based on its training to directly predict the mapping from grayscale photos with coloured dots to colour images. This method has the potential to save time for its users and can use the global histogram of a colour reference map to colour the grey map [23]. This research shows that the aforementioned techniques have given much thought to the problem of automatically colouring grayscale images. A difficulty, though, remains. There is a large theoretical research and practical application value for logistics firms [24] since, for example, no researcher has yet applied the CNN model to this subject.

3. **Proposed Work**

3.1 Dataset Preprocessing

We employ the 1.2 million high-resolution training pictures from the ImageNet ILSVRC2012 [12] dataset, which spans over 1000 categories; the hold-out validation set consists of 50,000 images for the proposed model. The evaluation split used, ctest10k, has 10,000 test pictures. A CNN-based picture colouring approach that can simultaneously detect and colour numerous image elements, such as foreground and background objects. In the first step, the semantic information is extracted using CNN. The colour from the selected region in the reference picture is applied to the selected region in the grayscale image based on the extracted semantic information. A sample of 10 images is tested for validating the effectiveness of the proposed model.

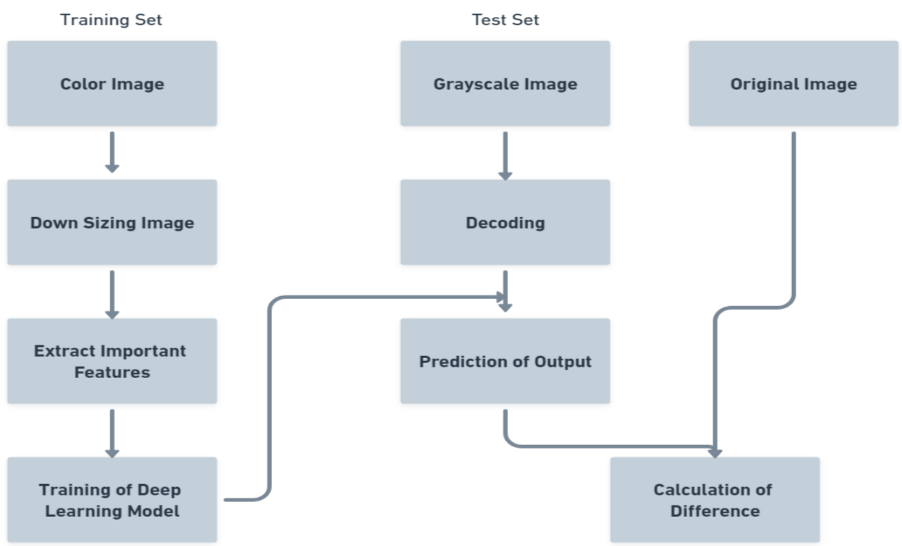


Figure.1 Proposed Block Diagram

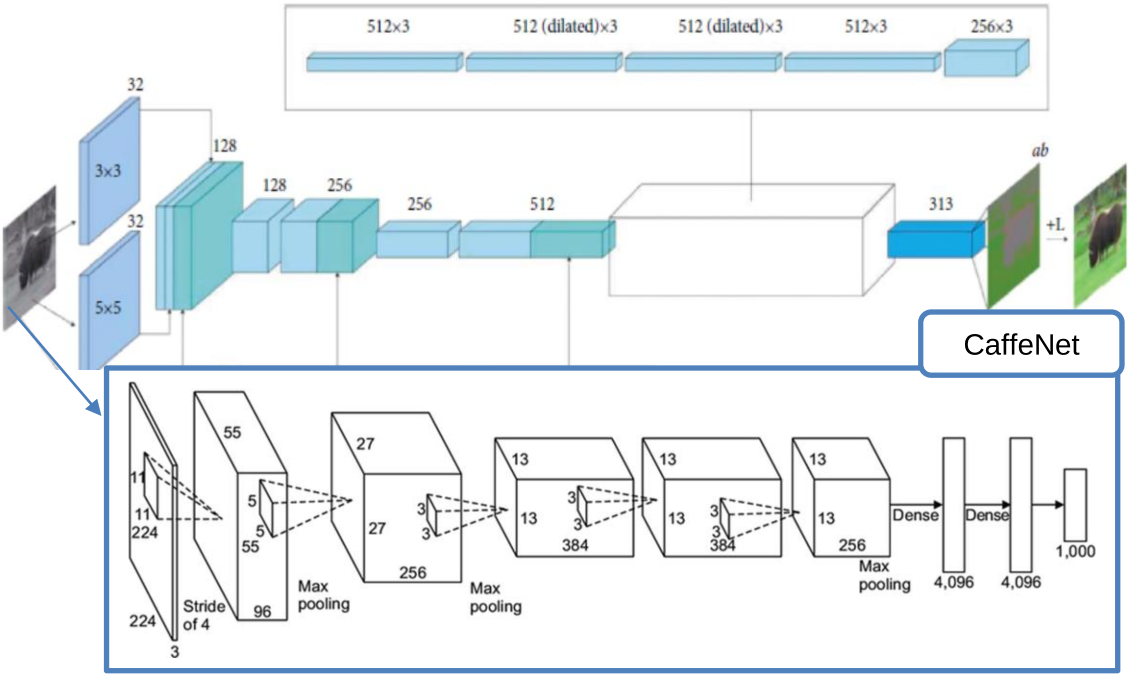


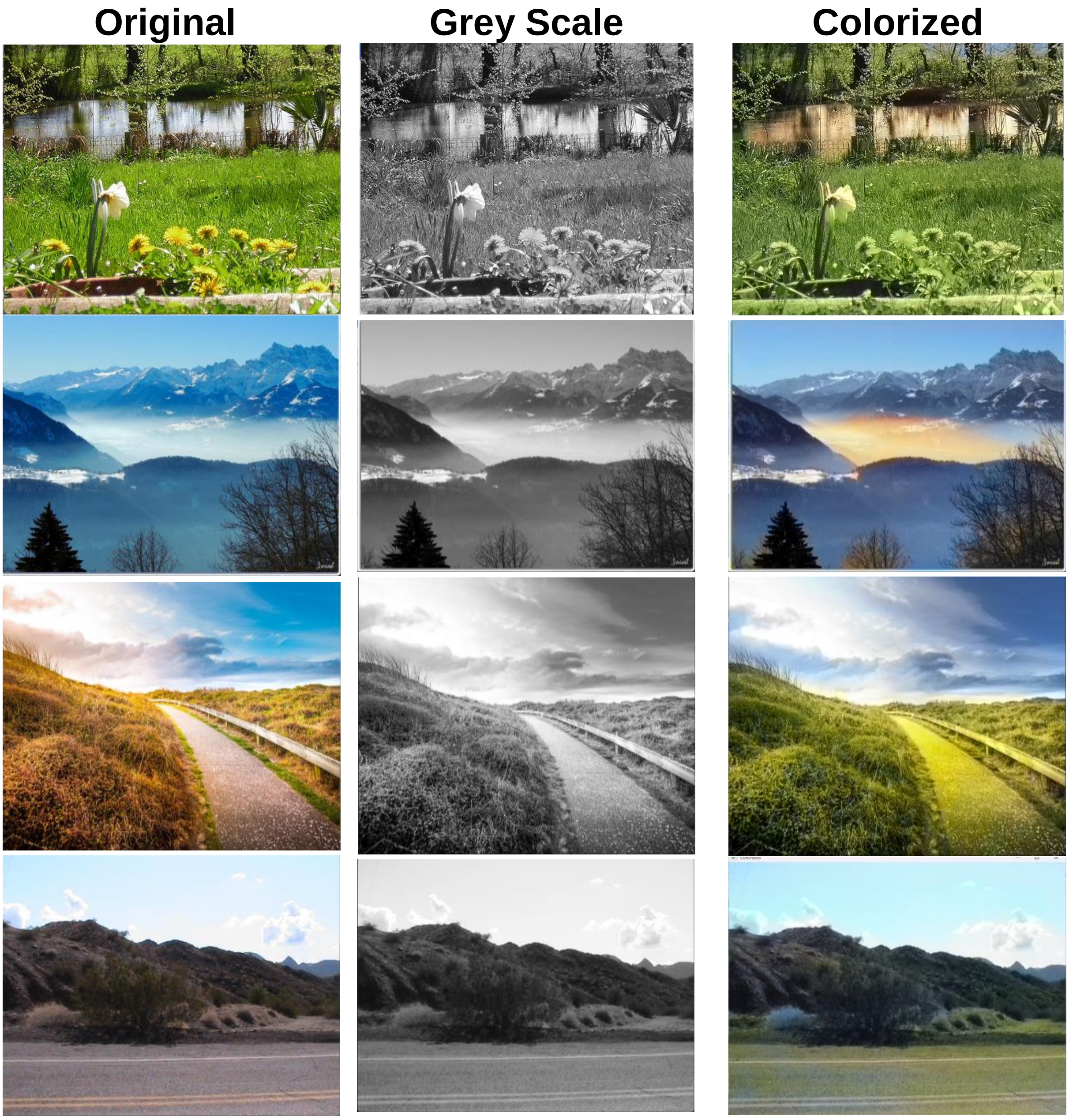
Figure.2 Proposed Modified version of CaffeNet

3.2 Proposed Model

Figure.1 shows the proposed block diagram and Figure.2 shows the modified version of the deep learning model. The basic convolution pooling network in AlexNet was passed down to the CaffeNet network, but instead of using huge convolution kernels, they switched to multiple convolution kernels with a size of 3×3, which could decrease the number of network parameters and boost the network depth. This is comparable to a more complex form of AlexNet; a deeper network is better suited to solving complex nonlinear issues. Even so, there are still a lot of CaffeNet network settings. A CaffeNet network typically has 500 parameters; hence, the model requires a large amount of storage. However, because of its superior feature extraction capabilities, it is ideally suited for feature extraction's supporting role in some image processing tasks. Differential networks refer to a structure that can be applied to any network model rather than a specific network. A connection mode called residual structure guards against network deterioration due to a hop connection. Additionally, even with the usage of the ReLU activation function, the phenomenon of gradient disappearance will happen as the number of network layers rises, whereas the residual structure uses the skip connection method in the network structure to address the aforementioned issues. We obtain the feature graph of the input feature graph after the convolution layer. Now, we intend to train the classifier using these feature graphs. The classifier can theoretically be trained using all of the collected feature graphs. We can apply the statistical method of aggregation to resolve this issue. For instance, we substitute the average of the image features for all of the original features, which is quicker and less prone to over-fitting than using all of the features. Pooling is the term used to describe this aggregation, and it is further separated into average pooling and maximum pooling. Here, we use the weight of each convolution kernel unit as an example of the average pooling approach. A bias unit is still added following each convolutional process.

4. **Results and Discussion**

Pytorch 1.8, Python 3.7, and CUDA for Deep Learning use of the Version 10.0 Data Set More than 50 types of outdoor settings are included here, from houses to cottages to landscapes to courtyards. The Epoch is 10, and the Epoch taught by a business line using the same data set is likewise 10. The experiment begins with an initial learning rate of 0.02 based on past results; the attenuation weight coefficient is 0.0001; the updated weight is 0.1; the updated weight attenuation system is 0.0002; the maximum number of iterations is 10,000; the epoch size is 600 times; and the random gradient descent method Batch is used. There are fifty distinct batch training possibilities.



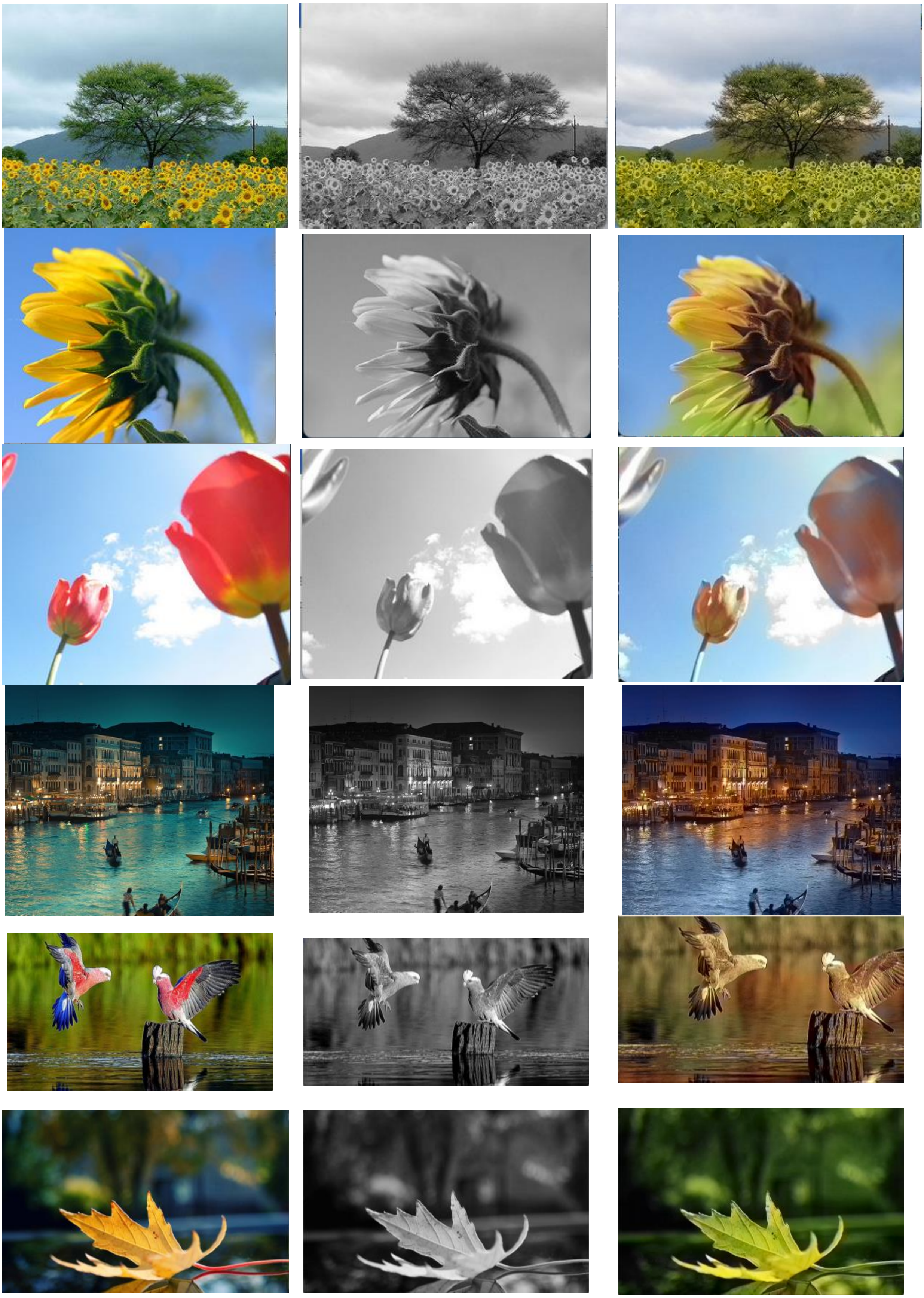


Figure.3 Qualitative analysis of the proposed technique

As can be seen in Figure 3, the colouring results may be both varied and helpful to the user's picture creation if they are drawn from a variety of categories. All three colouring techniques—foreground, back, and Poisson melting—are displayed in the final results. When one or more shells are present in the image, the user may choose a unique reference image for each marking and apply that colour to all ICONS in the image at once. One or more markings in the picture are coloured based on the supplied semantic map of the grayscale image. In addition, this paper's two colour labeling techniques have more stringent limitations thanks to the incorporation of semantic information as a strong constraint condition, allowing for a more desirable graphical colouring result to be achieved. The quality of the result of instance segmentation impacts the fusion effect, and the simple fusion effect is clearly dependent on the result of picture instance segmentation. However, there is still a shortage of edge processing, and existing instance segmentation technologies can only do a broad encirclement. This work employs the CNN method to blend the coloured foreground and backdrop into a single image, allowing for a seamless edge. The deeper the colour, the larger the end loss, but the result is still satisfactory because it is impossible to assess the final effect of colouring using a quantitative approach. The Pseudocolouring approach has a strong semantic features, vibrant colours, and good sky restoration, achieves a more natural colouring effect at the transition between foliage and sky and provides more precise building colouring. It also produces quite realistic ground colouring. This finding demonstrates that this paper has achieved the best image colouring effect.

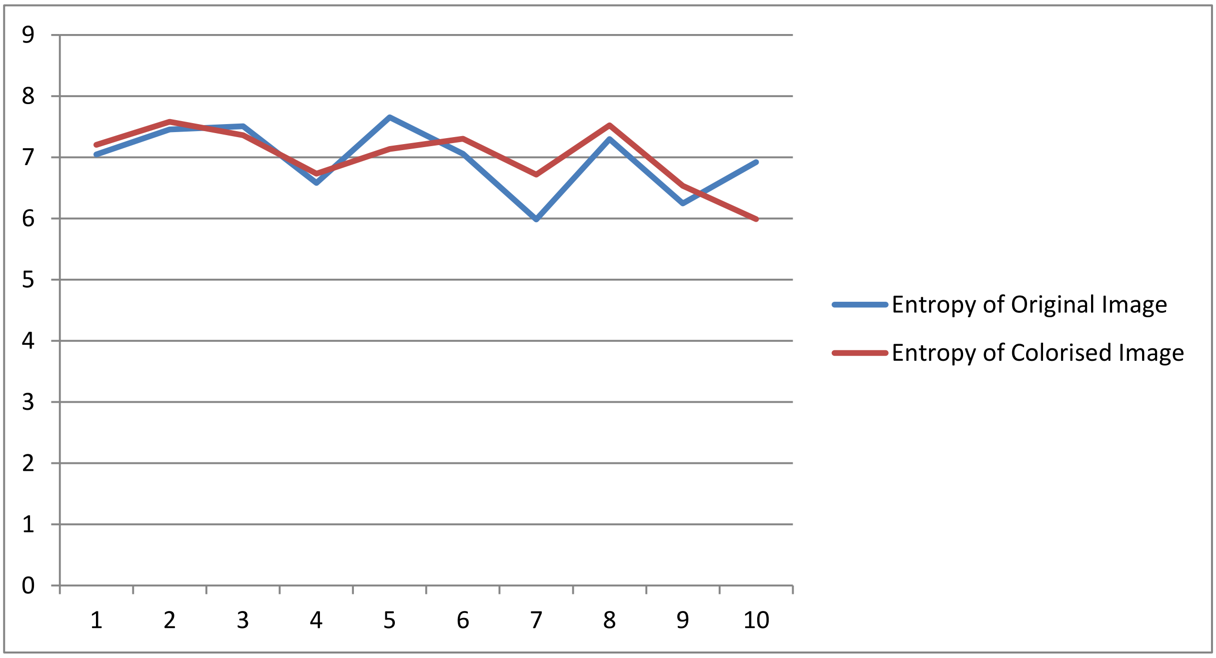


Figure. 4 Quantitative analysis- Entropy of the proposed technique

Entropy is a non-reference based criterion for judging the quality of an image. Entropy is not compared to any other entity or reference value; it is calculated to determine the amount of information present in the image. Entropy is determined here separately for the input and output images. Entropy, the average amount of information in the image, is used to standardise the information effectively. Entropy is the result of multiplying the base-2 logarithm of the probability values by the negative sum of the probability values. It does not have a set quantity or range because it is a non-reference based quality metric. From Figure.4, it is interpreted that there is an improvement in the entropy across all the images. This indicates the improvement in picture details, which is indicated by the higher output entropy when compared to input entropy.

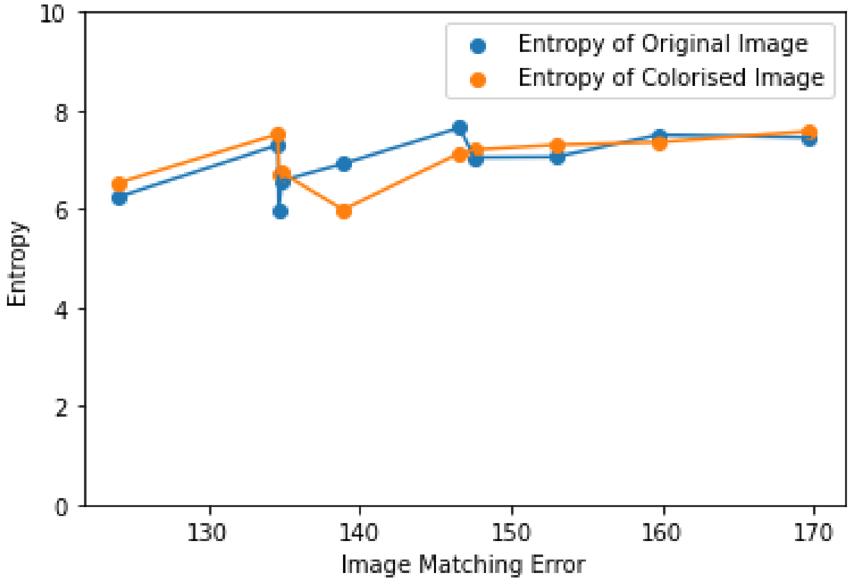


Figure. 5 Quantitative analysis- Image Matching error vs Entropy of the proposed technique

Figure 5 shows the image matching error variation for 10 randomly selected test images and how this error correlates with the entropies of the original and colourized images. It portrays the geometric differences between the two images. It is the difference between the corresponding pixels of the images taken into consideration. The graph from figure 5 makes it clear that, for the same image matching error, there isn't much of a difference between the entropies of the colourized and original images, indicating that the model has been successful in maintaining the randomness of the pixels and that the amount of pixel values that are different is relatively small. The figures above clearly show that this model can preserve the characteristics of the original image with the exception of a few off-kilter pixel values.

S. No.	Image Matching Error	Mean Square Error	Peak Signal to noise ratio (in dB)	Structural Similarity Index	Entropy of Original Image	Entropy of Colorized Image
1	147.539043	78.304817	29.192919	0.994720	7.047568	7.201979
2	169.699468	87.535491	28.708962	0.996647	7.454486	7.578864
3	159.705031	86.533978	28.758937	0.996671	7.504931	7.357714
4	134.850917	59.593126	30.378842	0.997361	6.577403	6.732229
5	146.593541	86.888746	28.741168	0.995250	7.650021	7.135920
6	152.944983	78.173954	29.200183	0.993878	7.057355	7.301948
7	134.645703	68.608051	29.767053	0.995173	5.984104	6.715244
8	134.575353	97.657457	28.233749	0.996271	7.297871	7.519803
9	124.035448	99.540906	28.150788	0.903903	6.245585	6.532116
10	138.892553	61.582066	30.236261	0.992173	6.920433	5.987555

Table 1: Performance metrics for the used images

A perceptual metric called the Structural Similarity Index (SSIM) calculates how much image quality is lost during data transmission errors or other processing stages like data compression. The full reference metric requires a reference image and a processed image from the same image capture. The

SSIM value is between 0 and 1. Generally, good quality reconstruction techniques have SSIM values of 0.97, 0.98, and 0.99. For the images used above, our model has an average SSIM of 0.9862047, which is good because it is very close to the value 1. The PSNR block calculates the peak signal-to-noise ratio (PSNR) between two images, which is expressed in decibels. This ratio is used to assess the quality of original and compressed images. With increasing PSNR, the quality of the compressed or rebuilt image gets better. A good PSNR is over 20 dB, and the average PSNR for the images used in our model is 29.1368862. Entropy, which aids in evaluating changes to an image, is a change in the randomness of a picture or image. The squared cumulative error between the original and compressed image is measured by the MSE. The error decreases as MSE decreases and also the average MSE of the model for used images is 80.4418592. The main distinction between image matching error and mean square error is that image matching error employs saturation to calculate the error, whereas MSE uses standard subtraction. The model employed has an average picture matching error of 144.34. The model performed exceptionally well overall for the images utilised, although performance metrics can change depending on the images.

5. Conclusion

In this proposed model, we suggest a novel CNN network for automated colouring of grey pictures, one that makes use of the color-based categorization of networks. The final colouring impact and loudness control are improved compared to the outcomes of other commercial approaches. CNN does exhaustive background colouring, but the target must be coloured using a colour reference graph. Iterative picture colouring and training image conversion are the two primary types of target colouring. Changing the colour of a grayscale image automatically is a promising area of study in the study of images. The proposed deep learning has the ability to automatically colourize grayscale images and it is possible to accomplish both automatic colouring and interactive colouring based on a reference image. But there are still some obstacles to overcome before deep learning can be used to fully complete the grey image automatic colouring task which include, but are not limited to, the lack of a unified and accurate method for evaluating grey image line contour recognition and the absence of a comprehensive professional platform dedicated to colouring grey images.

6. References

[1] X. Bi, W. Yao, Z. Zhang, S. Huang, J. Liu, and B. Chen, (2021), “Image steganography algorithm based on image colorization,” in Proceedings of the International conference on signal image processing and communication (ICSIPC 2021), Chengdu, Sichuan, China **vol. 11848**, Article ID 1184818.

[2] Bouida, A., Beladgham, M., Bassou, A., & Benyahia, I. (2020). Quality and texture analysis of biometric images compressed with second-generation wavelet transforms and SPIHT-Z encoder. *Indonesian Journal of Electrical Engineering and Computer Science*, ***19*****(3)**, 1325-1339.

[3] Liu, B., Gan, J., Wen, B., LiuFu, Y., & Gao, W. (2021). An automatic coloring method for ethnic costume sketches based on generative adversarial networks. *Applied Soft Computing****, 98,*** 106786.

[4] You, W. T., Sun, L. Y., Yang, Z. Y., & Yang, C. Y. (2019). Automatic advertising image color design incorporating a visual color analyzer. *Journal of Computer Languages*, ***55***, 100910.

[5] Kong, G., Tian, H., Duan, X., & Long, H. (2021). Adversarial edge-aware image colorization with semantic segmentation. *IEEE Access*, ***9,*** 28194-28203.

[6] Zare, M., Lari, K. B., Jampour, M., & Shamsinejad, P. (2019, March). Multi-GANs and its application for Pseudo-Coloring. In *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)* (pp. 1-6). IEEE.

[7] Zhang, J., Zhu, S., Liu, K., & Liu, X. (2022). UGSC-GAN: User-guided sketch colorization with deep convolution generative adversarial networks. *Computer Animation and Virtual Worlds*, ***33*(1),** e2032.

[8] Mohammed, B. N., & Ahmad, H. B. (2021). Advanced car-parking security platform using Arduino along with automatic license and number recognition. *Academic Journal of Nawroz University*, ***10*(1)**, 1-6.

[9] Chen, L., Han, J., & Tian, F. (2021). Colorization of fusion image of infrared and visible images based on parallel generative adversarial network approach. *Journal of Intelligent & Fuzzy Systems*, (Preprint), **vol. 41**, 1-10

[10] Jain, P., & Ghanekar, U. (2018). Robust watermarking technique for textured images. *Procedia Computer Science*, ***125***, 179-186

[11] Zhang, N., Qin, P., Zeng, J., & Song, Y. (2019). Image colorization algorithm based on dense neural network. *International Journal of Performability Engineering*, ***15*(1)**, 270.

[12] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

[13] Chen, Y., Luo, Y., Ding, Y., & Yu, B. (2018, July). Automatic colorization of images from Chinese black and white films based on CNN. In *2018 International Conference on Audio, Language and Image Processing (ICALIP)* (pp. 97-102). IEEE.

[14] Zeng, X., Tong, S., Lu, Y., Xu, L., & Huang, Z. (2020). Adaptive medical image deep color perception algorithm. *IEEE Access*, ***8***, 56559-56571.

[15] An, J., Kpeyiton, K. G., & Shi, Q. (2020). Grayscale images colorization with convolutional neural networks. *Soft Computing*, ***24*(7)**, 4751-4758..

[16] JWA, M., & KANG, M. (2021). Grayscale image colorization using a convolutional neural network. *Journal of the Korean Society for Industrial and Applied Mathematics*, ***25*(2),** 26-38.

[17] Farid, M. S., Lucenteforte, M., & Grangetto, M. (2018). Evaluating virtual image quality using the side-views information fusion and depth maps. *Information Fusion*, ***43***, 47-56.

[18] Berger, D. R., Seung, H. S., & Lichtman, J. W. (2018). VAST (volume annotation and segmentation tool): efficient manual and semi-automatic labeling of large 3D image stacks. *Frontiers in neural circuits*, ***12***, 88.

[19] Thakur, G. K., Priya, B., & Mishra, R. K. (2019). An efficient coloring algorithm for time detraction of sign image segmentation based on fuzzy graph theory. *Journal of Applied Security Research*, ***14*(2)**, 210-226.

[20] Oladi, M., Ghazilou, A., Rouzbehani, S., Polgardani, N. Z., Kor, K., & Ershadifar, H. (2022). Photographic application of the Coral Health Chart in turbid environments: The efficiency of image enhancement and restoration methods. *Journal of Experimental Marine Biology and Ecology*, ***547*,** 151676.

[21] Qiao, T., Shi, R., Luo, X., Xu, M., Zheng, N., & Wu, Y. (2018). Statistical model-based detector via texture weight map: Application in re-sampling authentication. *IEEE Transactions on Multimedia*, ***21*(5),** 1077-1092.

[22] Ouyang, D., Furuta, R., Shimizu, Y., Taniguchi, Y., Hinami, R., & Ishiwatari, S. (2021). Interactive Manga Colorization with Fast Flat Coloring. In *SIGGRAPH Asia 2021 Posters* (pp. 1-2).

[23] Huang, S., Jin, X., Jiang, Q., Li, J., Lee, S. J., Wang, P., & Yao, S. (2021). A fully-automatic image colorization scheme using improved CycleGAN with skip connections. *Multimedia Tools and Applications*, ***80*(17),** 26465-26492.