A CNN Based Deep Learning Model for Black and White Image Colorization

Apurva Anand[1], Srungavarapu Suhaas[1], Utkarsh Anand[1], and Anand Bihari[1,*]

A R T I C L E I N F O

Article history:
Received: 31 July 2023
Revised: 15 October 2023
Accepted: 7 November 2023
Online: 20 June 2024

Keywords:
Colorization
Graphics vision
CNN
RGB
Black and white
Image processing
LAB color

A B S T R A C T

Colorization is the process of transforming black and white images into aesthetically appealing color images. The main goal is to persuade the viewer that the outcome is real. Automatic conversion has become a complex area that combines machine learning, deep learning, and art. Image colorization is one method of adding style to a photograph or combining styles. Image colorization can also be used to add color to photographs that were previously black and white. This can be utilized to make an educated assumption about the picture's context and bridge the gap between the past and the present. Unlike the previous techniques, the goal of this work is to develop a high-quality fully-automatic colorization system to color the black and white images. By training our model on ImageNet images, we were able to get it to produce shots with realistic hues with 85.47% accuracy. Also the outcome of the proposed model is validated with the peer user by online survey of the colored images.

1. INTRODUCTION

The process of adding a color to each pixel of a target grayscale image is referred to as image colorization. Colorization approaches are broadly classified as scribble-based colorization [1, 2, 3, 4, 5] and example-based colorization [6, 7, 8, 9]. Scribble-based approaches often need significant user effort to generate significant scribbles on the target grayscale photos. It is so time-consuming, especially for a novice user, to colorize a grayscale image with fine-scale features.

In Example based method, the Color information is frequently transferred from a comparable reference image to the target grayscale image. Finding an appropriate reference image is one of the major challenges for a user. To simplify the problem by using picture data from the Internet and provide filtering algorithms for selecting appropriate reference photos. But both the measure required additional constraints [8, 10]. An identical Internet object is required for exact per-pixel registration between the reference photos and the grayscale target image. As a result, it is confined to objects having a fixed shape like landmarks. In some cases, if the image contains text and other material with the image, then user is required to give a semantic text label as well as segmentation information for the foreground item [8,10]. In manual segmentation, it is very hard to get the original grayscale image into the colored image when image contains several complex objects. The quality of the colored image is completely depending on the selected reference image and the segmentation information provided by the user. One of the most significant uses of the grayscale image matting approach is to combine it with color transferring techniques to produce object-based colorization, which colorizers' objects in the same picture individually. Grayscale picture colorization has uses in black and white photo editing, colorization of vintage films, and scientific drawings. Colorization may significantly improve the visual appeal of grayscale photographs and perceptually improve scientific representations.

Welsh et al. [11] suggested a grayscale picture colorization approach that performs well for natural and scientific illustrative images. Welsh et al.'s technique, in general, works best on scenarios where the image is separated into different brightness clusters or when each region has distinct textures. Their present approach, however, does not function well with human faces. These methods are too complex and require human involvement or less impact of colorization of image. As a result, it is required a pure automatic colorization method to address these limitations. One of the solutions could be the conversion of RGB color to the LAB color space.

Influenced by the current achievements of deep learning methods in image processing, we employ CNN

[1]*School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tanilnadu 632014, India.*
**Corresponding author:* Anand Bihari; Email: anand.bihari@vit.ac.in, csanandk@gmail.com.

(Convolutional Neural Network) to investigate colorization problems in photographs. However, the purpose of this research is to create a realistic colorization that may potentially trick a human observer, rather than to retrieve the actual ground truth color. As a result, our goal becomes much more attainable: model adequate grayscale semantics and textures have statistical dependencies.

In this paper, an introduction to the image colorization and its importance is presented in Introduction section. Related work section presented the previous works done by the eminent scholars. The proposed methods section presents the data set details, proposed methods and the experiment details. Result and discussion section presents results analysis and discussion and section Conclusion concludes the work.

2. RELATED WORK

We cover the key components of prior work in this section.

Gupta et al. (2012) [12] presented an example-based method for colorizing a grey image is presented in this research. In this research author has suggested that the user just give the color image as reference that is semantically comparable to the output image. They utilize super-pixel resolution for feature extraction from these photos, which they then use to guide colorization. The colorization procedure is sped considerably by their usage of a super-pixel representation. They create an image space voting system that uses information from adjacent super-pixels to identify and fix inaccurate color assignments, further ensuring the coherence of these first color designations in space. For example, while super-pixel representation leads to greater spatial coherence, it can be inaccurate at object edges or thin image structures while colorizing. Bleeding artifacts at object boundaries could result as an outcome of this. Second, image segments generated during color reassignment in densely textured areas are frequently very small. Because these segments have fewer super-pixels than larger segments, the image voting step's robustness is lowered. As a reason, super-pixel voting for colors within these segments becomes less accurate. Finally, this technique relies on a color exemplar that is semantically equivalent to the grey image. Consequently, when suitable color exemplars are absent, our method may fail.

Cheng et al. (2015) [13] discussed the image colorization and proposed a novel, fully automated colorization system that uses deep neural networks to eliminate human labor and reliance on color examples. In this research author has supplied some new semantic feature including DAISY and Patch to the neural network as inputs and serve as helpful but discriminative characteristics. It is, however, dependent on machine learning techniques and comes with its own set of limitations. It's designed to be trained on a large reference photo collection, such as a collection of all imaginable items. In actuality, this is impossible to achieve. That means, during conversion from color to grayscale, the several pixels were lost.

Deshpande et al. (2015) [14] discussed large-scale image colorization. This paper predicts colorization depending on a Learch objective [24] that is actively learned. They indicate that the technique provides spatially coherent colorizations that are visually appealing and convincing when paired with histogram correction. The best performance is obtained when the information about the scene is provided. This completely automated approach achieves near-optimal results.

Iizuka et al. (2016) [15] proposed a new architecture for colorizing grayscale photos that combine international and domestic information. This method uses deep neural networks and can colorize images without human interaction. The model is trained for scene identification with a combined colorization and transmission line losses that allows it to recognize colors and adjust according to the image's aspect; for example, the sky color in a setting sun image differs from the sky color in a daylight image. The best part of the suggested architecture is to user can process any resolution image with affecting the quality of the image. We can also do style transfer or color a picture using the background of another, using the same approach. The method's prime requirement is that it is data-driven, meaning it can only colorize photos with identical qualities to those in the training set. To mitigate this, they test the model with a huge broad selection of interior and outdoor scene photos. Images created by humans, on the other hand, are not included. It would be essential to train a new model for the new photographs if they wanted to analyze considerably different types of images.

Pahal and Sehrawat (2016) [16] presents a deep convolution network-based method for reliably colourizing black-and-white photographic images without requiring direct human interaction. As our research has shown, this method also has the benefit and promise of employing CNN model to colorize black and white photos. We've demonstrated that structuring the problem as a classification task can result in colourized photographs that are arguably considerably more aesthetically beautiful than those generated by a basic regression-based algorithm, indicating that it has a lot of room for improvement.

Guadarrama et al. (2017)[17] PixColor generates a wide range of colorizations, and in a crowd-sourced human evaluation, the model's results outperformed other published methods on average. They get over Pixel CNN's sluggish inference difficulty by sampling only low-resolution color channels and then improving the output with a regular image-to-image CNN. In preliminary studies, the likely score (as determined by the Pixel CNN model) was used to select the best sample., so there was no good correlation with human judgment.

Liu et al. (2018) [18] designed a colorization technique based on examples resistant to lighting changes between grayscale targets and color reference images. The approach manages this by executing color transfer in an illumination-independent area with few shadows and highlights. It starts by gathering several color references from the web to create an illumination-independent intrinsic reflectance picture of the sample image. The example photographs found through a web search could have been taken from various viewpoints, under various lighting conditions, and potentially different. The target grayscale image is then disassembled into its fundamental reflectance and illumination components using grayscale replicas of these referenced images. To obtain the final output, color is transferred by converting the color reflectance image to the grayscale reflectance image and afterward relighting using the target image's illumination component. Intrinsic colorization implies that all images are lit with white light. The intrinsic image decomposition must function for it to work. The approach needs enough reference photos from the web to supply enough registerable images with varied illuminations. Furthermore, this method is limited to colorizing static scenes using photographs of the same scene taken from similar angles. It's possible that data in the target image that differs from the reference image, or items from prior shots that have vanished, won't be colorized correctly.

Žeger, Ivana, et al (2021) [19] represents an overview and assessment of grayscale image colorization methods and techniques used for natural images. In this paper author has developed a colorization using deep learning methods.

Form the above discussion; it is found that most of the research required user intervention that reflects the quality of the image. Our proposed research will address this issue.

3. PROPOSED METHOD

This method generates a realistic color representation of a grayscale photograph. Due to the obvious limitations of this problem, previous solutions either depended heavily on human intervention or resulted in desaturated colorizations. We have developed a fully automated method for creating colorful and realistic colorizations to overcome these limitations. In this model, we have used “ImageNet Dataset” [20] to train the model, a picture database arranged as per the WordNet tree [22], with hundreds of thousands of photos depicting each node in the hierarchy. The model was trained with over a million color images and used as a feed-forward phase in a CNN [29, 30] during testing. For the photographs in the dataset, the RGB color scheme will be converted to “LAB color space.”

3.1 Datasets

ImageNet [20] is a huge database of annotated images that is utilized in computer vision and image processing. The dataset was produced with the goal of providing a resource for researchers and developers working on improving computer vision technology.

The collection contains somewhat more than 14 million pictures, somewhat more than 21 thousand sets or classes (synsets), and somewhat more than 1 million pictures with bounding box annotations, according to data on the ImageNet webpage.

Grayscale photos were filtered out of the training, validating, and testing sets because they were in the ImageNet dataset. For this experiment a total of 20,000 grayscale image, 100 legacy images and 100 man-made has been used. This dataset is further splited into 70 : 20 : 10 for training, testing and validation.

3.3 Proposed Model Description:

The proposed method utilizes the architecture given in ref [21]. The architecture is designed with the help of CNN, which converts a grayscale source to a distribution of quantized color value outputs. Here we have changed the input parameter of the given architecture to LAB color and Softmax activation function [23]. The proposed model algorithm is given in algorithm 1.

Algorithm 1

1. Input dataset and libraries
2. Preprocessing of data (extraction of data from image)
3. Convert RGB to LAB color space
4. Building the model using CNN
5. Use the L channel as the input, and then train it to anticipate the ab channels.
6. Combine the predicted (a, b) Probability distribution channels (using defined model) with the input L channel.
7. Return the Lab image to RGB image.
8. End

In this method we have used 8 convolutional layers that repeat a block of 2-3 layers. In this model we have not used any pooling layer. The up and down sampling has been used between the blocks to get correct resolution in all changes. The detailed parameter of the proposed method is given in table 1.

Table 1: Proposed model parameter description

	SRO	CN	STR	KD	STRa	DE
Data	224	3	-		-	-
Conv1.1	224	64	1	1	1	1
Conv1.2	112	64	2	1	1	1
Conv2.1	112	128	1	1	2	2
Conv2.2	56	128	2	1	2	2
Conv3.1	56	256	1	1	4	4
Conv3.2	28	256	1	1	4	4
Conv3.3	28	256	2	1	4	4
Conv4.1	28	512	1	1	8	8

Conv4.2	28	512	1	1	8	8
Conv4.3	28	512	1	1	8	8
Conv5.1	28	512	1	2	8	16
Conv5.2	28	512	1	2	8	16
Conv5.3	28	512	1	2	8	16
Conv6.1	28	512	1	1	8	16
Conv6.2	28	512	1	1	8	16
Conb6.3	28	512	1	1	8	16
Conv7.1	28	256	1	1	8	8
Conv7.2	28	256	1	1	8	8
Conb7.3	28	256	1	1	8	8
Conv8.1	56	128	1	1	4	4
Conv8.2	56	128	1	1	4	4
Conb8.3	56	128	1	1	4	4

Here **SRO** represents the spatial resolution of output, **CN** represents number of channels of output, **STR** represents the computation stride, **KD** represents kernel dilation, **STRa** represents the accumulated stride across all preceding layers and **DE** represents the effective dilation of the layer with respect to the input.

4. RESULT AND DISCUSSION

In this section we are going to present the outcome of our research with validation. The proposed model has been trained and tested with the ImageNet dataset. That is publically available dataset. At the time of training and testing a total of 10k image data has been used. In this research we have also tested the some of the legacy black and white image. Further we have made validated the proposed method with the survey with independent user. The sample output of the image Net dataset is given in Table 2.

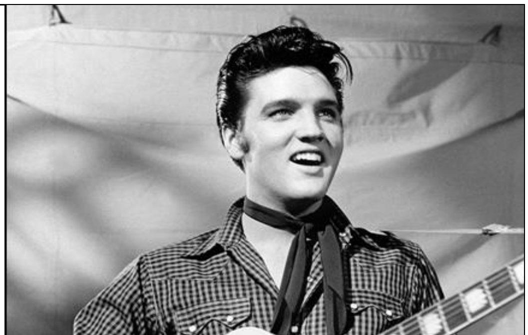
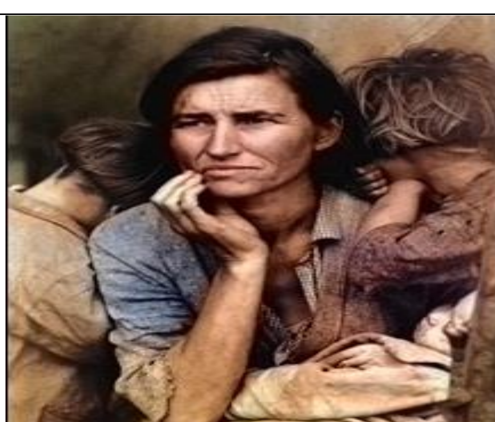
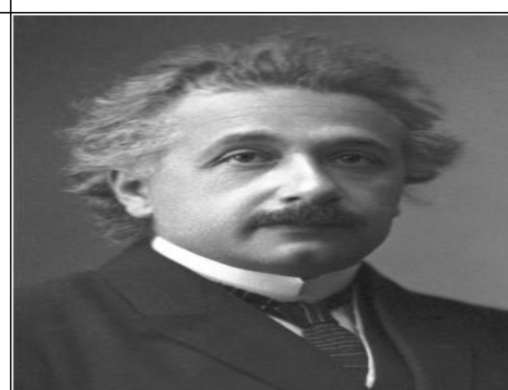
Table 2: The comparative result of the proposed model with original and ground truth image

Input-1	
Proposed Model Output	
Real Image	
Input-2	
Proposed Model Output	
Real Image	

Input-3	
Proposed Model Output	
Real Image	
Input-4	
Proposed Model output	
Real Image	
Input - 5	
Proposed Model Output	
Real Image	

From the table 2 we can conclude that the proposed method colorization is almost near to the ground truth images. We have tested the model accuracy and achieved a total of 85.47% accuracy. To validate the proposed model, we have colorized some legacy black & white image (source ImageNet dataset) and conducted an online survey with local pears and got 65% confidence. The legacy image colorization output is shown in Table 3.

Table 3: Proposed model output on Legacy Black and White images

Input-1	
Proposed Model Output	
Input -2	
Proposed Model Output	
Input -3	
Proposed Model Output	
Input - 4	

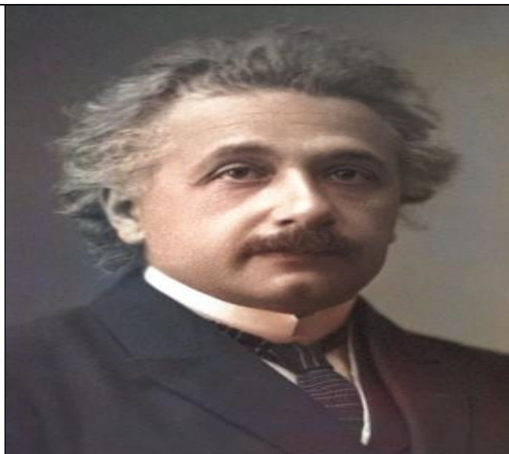
Proposed Method Output	
Input - 5	
Proposed Model Output	

Table 3 shows the outcome of the legacy black and white images. For further validation we have tested our proposed model with man-made objects (Source-ImageNet) and found most of the cases we did not reached to the original quality. The man-made object colorization is shown in table 4.

Table 4: Proposed model outcome is based on man-made images

Input-1	
Proposed model output	
Real Image	
Input-2	

Proposed Model Output	
Real Image	
Input - 3	
Proposed Method Output	
Real Image	
Input - 4	
Proposed Method Output	
Real Image	

From the above tables table 1, 2 and 3, we can conclude that the proposed method is quite suitable for the natural image and legacy black and white images, but it is not suitable for the man-made objects.

To validate the suggested approach, we compared it to the state-of-the-art methods that employed a wide range of photos. Table 5 displays the comparative result. Based on Table 5, we may infer that the suggested strategy outperforms the others.

Table 5: Comparative result of the proposed method with state of the art methods

Reference	Method used	Overall Accuracy
Zhou et al. (2014) [25]	ImageNet CNN feature + SVM	68.5%
Wang et al. (2015) [26]	Places205-VGGNet-11	82.3%
Iizuka et al. (2016) [27]	CNN	80.6%
Zhao et al. (2020) [28]	Deeplab-ResNet101	66.9%
Proposed model	**CNN+LAB color**	**85.47%**

5. CONCLUSIONS

This research describes a unique, fully automatic colorization approach that use CNN to reduce human effort and enhance the image colorization as well as acceptability of the colorization. Image colorization is one method of adding style to a photograph or combining styles. Image colorization can also be used to add color to photographs that were previously black and white. In this proposed method the RGB color is converted into LAB color and then training and testing of the model is done. By training our model on ImageNet images dataset, we were able to get it to produce shots with realistic hues. Our method achieves 85.47% accuracy, requires fewer hand-tuned parameters and elements, and has been validated on a larger and more diversified set of test samples. Further our proposed method is implemented for legacy black and white images and for man-made images. The proposed model is quite suitable for natural image and legacy black and white image but not much suitable to colorize the man-made images. This research is limited to implementation of gray scale images, but several man-made images and black and white videos are also available required colorization.

REFERENCES

[1] Huang, Y. C., Tung, Y. S., Chen, J. C., Wang, S. W., & Wu, J. L. (2005, November). An adaptive edge detection based colorization algorithm and its applications. In Proceedings of the 13th annual ACM international conference on Multimedia (pp. 351-354).

[2] Levin, A., Lischinski, D., & Weiss, Y. (2004). Colorization using optimization. In ACM SIGGRAPH 2004 Papers (pp. 689-694).

[3] Luan, Q., Wen, F., Cohen-Or, D., Liang, L., Xu, Y. Q., & Shum, H. Y. (2007, June). Natural image colorization. In Proceedings of the 18th Eurographics conference on Rendering Techniques (pp. 309-320).

[4] Qu, Y., Wong, T. T., & Heng, P. A. (2006). Manga colorization. ACM Transactions on Graphics (ToG), 25(3), 1214-1220.

[5] Yatziv, L., & Sapiro, G. (2006). Fast image and video colorization using chrominance blending. IEEE transactions on image processing, 15(5), 1120-1129.

[6] Charpiat, G., Hofmann, M., & Schölkopf, B. (2008). Automatic image colorization via multimodal predictions. In Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III 10 (pp. 126-139). Springer Berlin Heidelberg.

[7] Chia, A. Y. S., Zhuo, S., Gupta, R. K., Tai, Y. W., Cho, S. Y., Tan, P., & Lin, S. (2011). Semantic colorization with internet images. ACM Transactions on Graphics (ToG), 30(6), 1-8.

[8] Gupta, R. K., Chia, A. Y. S., Rajan, D., Ng, E. S., & Zhiyong, H. (2012, October). Image colorization using similar images. In Proceedings of the 20th ACM international conference on Multimedia (pp. 369-378).

[9] Ironi, R., Cohen-Or, D., & Lischinski, D. (2005). Colorization by Example. Rendering techniques, 29, 201-210.

[10] Liu, X., Wan, L., Qu, Y., Wong, T. T., Lin, S., Leung, C. S., & Heng, P. A. (2008). Intrinsic colorization. In ACM SIGGRAPH Asia 2008 papers (pp. 1-9).

[11] Welsh, T., Ashikhmin, M., & Mueller, K. (2002, July). Transferring color to greyscale images. In Proceedings of the 29th annual conference on Computer graphics and interactive techniques (pp. 277-280).

[12] Gupta, R. K., Chia, A. Y. S., Rajan, D., Ng, E. S., & Zhiyong, H. (2012, October). Image colorization using similar images. In Proceedings of the 20th ACM international conference on Multimedia (pp. 369-378).

[13] Cheng, Z., Yang, Q., & Sheng, B. (2015). Deep colorization. In Proceedings of the IEEE international conference on computer vision (pp. 415-423).

[14] Deshpande, A., Rock, J., & Forsyth, D. (2015). Learning large-scale automatic image colorization. In Proceedings of the IEEE international conference on computer vision (pp. 567-575).

[15] Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2016). 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 (ToG), 35(4), 1-11.

[16] Pahal, S., & Sehrawat, P. (2019, December). Image colorization with deep convolutional neural networks. In International Conference on Advanced Communication and Computational Technology (pp. 45-56). Singapore: Springer Nature Singapore.

[17] Guadarrama, S., Dahl, R., Bieber, D., Norouzi, M., Shlens, J., & Murphy, K. (2017). Pixcolor: Pixel recursive colorization. arXiv preprint arXiv:1705.07208.

[18] Liu, X., Wan, L., Qu, Y., Wong, T. T., Lin, S., Leung, C. S., & Heng, P. A. (2008). Intrinsic colorization. In ACM SIGGRAPH Asia 2008 papers (pp. 1-9).

[19] Žeger, I., Grgic, S., Vuković, J., & Šišul, G. (2021). Grayscale image colorization methods: Overview and evaluation. IEEE Access, 9, 113326-113346.

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

[21] Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14 (pp. 649-666). Springer International Publishing.

[22] Fellbaum, C. (Ed.). (1998). WordNet: An electronic lexical database. MIT press.
[23] Goodfellow, I., Bengio, Y., & Courville, A. (2018). Softmax units for multinoulli output distributions. Deep Learning.
[24] He, J., Sivanrupan, G., Tsankov, P., & Vechev, M. (2021, November). Learning to explore paths for symbolic execution. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (pp. 2526-2540).
[25] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. Advances in neural information processing systems, 27.
[26] Wang, L., Guo, S., Huang, W., & Qiao, Y. (2015). Places205-vggnet models for scene recognition. arXiv preprint arXiv:1508.01667.
[27] Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2016). 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 (ToG), 35(4), 1-11.
[28] Zhao, J., Han, J., Shao, L., & Snoek, C. G. (2020). Pixelated semantic colorization. International Journal of Computer Vision, 128, 818-834.
[29] Uoc, N. Q., Duong, N. T., Son, L. A., & Thanh, B. D. A (2022) Novel Automatic Detecting System for Cucumber Disease Based on the Convolution Neural Network Algorithm. GMSARN International Journal 16 (2022) 295-301
[30] Paritosh Bhushan, Md Shah Fahad, Sanjay Agrawal, Paritosh Tripathi, Praveen Mishra, and Akshay Deepak (2023) A Self-Attention Based Hybrid CNN-LSTM for Speaker-Independent Speech Emotion Recognition. GMSARN International Journal 17 (2023) 429-435.