Signal, Image and Video Processing
https://doi.org/10.1007/s11760-021-02047-5

ORIGINAL PAPER

Removal of spatial inconsistencies in automated image colorization using parameter-free clustering and convolutional neural networks

Navjot Singh[1] · Garvit Gupta[2] · Anubhav Singh[2] · Anshul Kishore[2] · Kumud Kumar[2] · Deepak Bharti[2]

Received: 3 November 2020 / Revised: 4 August 2021 / Accepted: 30 September 2021

Abstract
Given a grayscale photograph, this paper tackles the issue of fantasizing about a conceivable shading rendition of the photograph. This issue is underconstrained, so past methodologies have either depended on significant human cooperation or came about in desaturated colorizations. We propose a wholly automated approach that produces lively furthermore, sensible colorizations. This approach utilizes the combination of CNN-based colorization and parameter-free k-means clustering to identify the color spills, so that the image can be recolored to produce color images that are aesthetically more pleasing and plausible than the images produced by the state-of-the-art methods. The performance of the proposed model is evaluated in terms of mean square error and structural similarity and found to be superior to the related works.

Keywords Colorization · Convolutional neural networks · k-means clustering · Davies–Bouldin index · Computer vision

1 Introduction

Colorization [1, 2] adds colors to black and white images to make them look aesthetically pleasing. But this task has been identified to be challenging and laborious for a long. Despite several works on it, it still requires human efforts to achieve the plausible results. When we first encounter a black and white image, it seems that most of the information regarding the colors are permanently lost. If we inspect closely, we find that the image itself reveals sufficient hints and clues to how it should be colored. Like, in an image, grass/trees are usually green, water or sky is majorly blue, and fruits have their respective dominant colors. This makes it easier to identify the colors of various regions in the image. But this may not be the case for every image since different types of objects or scenes may be present in the image, which may acquire a large variety of colors, and each one of them is plausible. For instance, a T-shirt may be of color, say black, blue or red, or a rose may be of red, yellow, white, or pink color. This problem makes it difficult to achieve a colorization which is as accurately colored as the ground truth. But it also gives leverage to produce a plausible colorization such that a human could not distinguish between the produced output and the ground truth, thus, making our task achievable. Colorization procedures locate a few viable applications, for example, coloring old motion pictures or photographs and adjusting shading in inheritance pictures. This is an under-obliged issue because numerous hues can be allotted to a pixel with known greyscale intensities. Along these lines, human mediation frequently assumes a significant function in the colorization cycle as there is no great answer to the colorization issue.

In the literature, researchers have proposed techniques involving interactive colorization, colorization through user interaction, colorization through web interaction, and colorization through deep learning. It is detailed in the related works section. To discuss a few, Zhang et al. [2, 3] proposed

✉ Navjot Singh
navjot@iiita.ac.in

Garvit Gupta
garvitgupta58@gmail.com

Anubhav Singh
anubhavshipra@gmail.com

Anshul Kishore
anshulkishore310@gmail.com

Kumud Kumar
kumud.kumar85@gmail.com

Deepak Bharti
deepakbharti@mnnit.ac.in

1 Indian Institute of Information Technology Allahabad, Prayagraj 211015, India

2 Motilal Nehru National Institute of Technology Allahabad, Prayagraj 211004, India

Published online: 29 January 2022

two well-known models. The first being an automated system of image colorization using a deep convolutional neural network (CNN), and the second being a user-interactive system of colorization also using CNN. Both techniques treat the task of image colorization as a classification problem. They are trained on the ImageNet Dataset, which comprises as many as 14 M images. The following problems arise in these models.

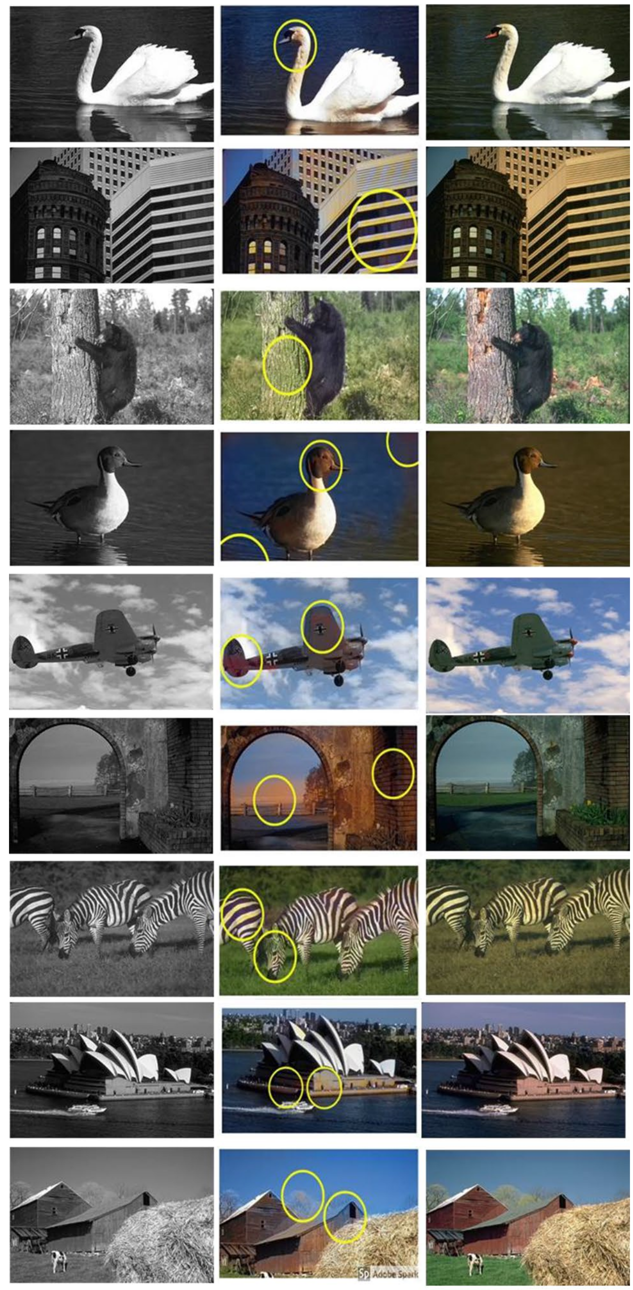
- The work of Zhang et al. [2] produces colorization that is spatially inconsistent, i.e., the colors are spilled around the edges, as shown in Fig. 1. This makes the produced image implausible and, thus, very different from the ground truth.
- The work by Zhang et al. [3] produces spatially consistent colorizations that are plausible and aesthetically pleasing, but it is an interactive system, as shown in Fig. 2, i.e., requiring user assistance, which is a challenge to make the system entirely automated. The user must mark the different regions in the image with the possible colors, and the system then produces the user-guided colorization.

These challenges posed by both the works give rise to the approach adopted by us to produce better results. In our work, we attempted to remove these spatial inconsistencies by automatically providing suitable colors. The contributions of the paper are as follows.

- Removal of spatial inconsistencies that remains after coloring.
- A fully automated system for colorization.
- Selection of suitable colors by employing a parameter-free clustering technique.

The proposed model includes three stages. In the first stage, a pre-trained colorization model proposed by Zhang et al. [2] is conveyed to shading a grayscale picture. At that point, the hues spilled along the edges are distinguished in the subsequent stage by picking an ideal number of shading groups. Finally, in the third stage, colors picked are passed to the user-guided picture colorization, as proposed by Zhang et al. [3], to improve the aesthetics of the picture.

The rest of the manuscript is organized as follows. Related works are discussed in Sect. 2. The proposed model is described in Sect. 3. The experimental setup and results are elaborated in Sect. 4. Finally, the conclusion and future work is stated in Sect. 5.



(a) (b) (c)

Fig. 1 **a** Original image, **b** spatial inconsistencies in the output of the model proposed by Zhang et al. [2], and **c** ground truth

2 Related works

This section is divided into four categories, namely (1) interactive colorization, (2) example-based colorization via user Interaction, (3) example-based colorization via web interaction, and (4) example-based colorization via deep learning. These are discussed in detail underneath.

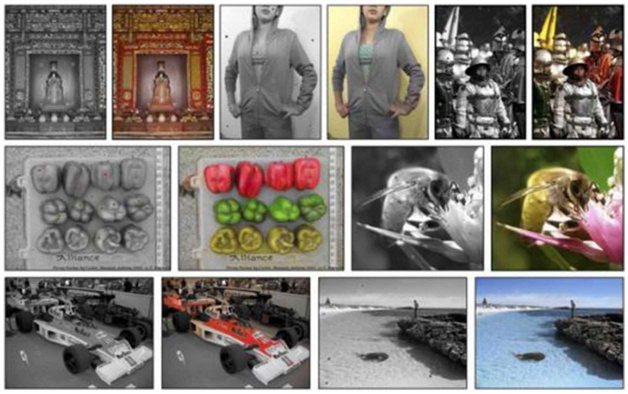


Fig. 2 Input and output of the model proposed by Zhang et al. [3]

Interactive colorization—Levin et al. [4] proposed a basic yet compelling methodology that requires the client to give vivid scribbles physically on the greyscale target picture as essential pointers and spread those shading scribbles to the whole picture consequently. The shading data on the scribbles is then engendered to the remainder of the target picture utilizing least-square streamlining. Huang et al. [5] built up a versatile edge location calculation, which improved the proliferation strategy by decreasing shading mixing at edges. Yatziv et al. [6] colorized the pixels utilizing a weighted blend of client jots. Qu et al. [7] and Luan et al. [8] used the surface element to lessen the required scribbles.

Example-based colorization via *user interaction*—Charpiat et al. [9] proposed a worldwide advancement calculation by assessing contingent probabilities over surface highlights and authorized perfection utilizing diagram cuts. Gupta et al. [10] built up a colorization technique given superpixel to improve the spatial coherency. The colorization nature of these techniques depends vigorously on the model image(s) given by the client. Notwithstanding, finding a good reference picture is a difficult undertaking because of the absence of standard models.

Example-based colorization via *web interaction*—Liu et al. [11] registered an inherent picture utilizing many comparative reference pictures gathered from the Internet. This strategy is hearty to brightening contrast between the objective and reference pictures; however, it cannot colorize the dynamic objects like vehicles, among the reference and target pictures, since these objects are barred from the calculation of the characteristic picture. Subsequently, it is restricted to static scenes and unbending formed scenes like famous milestones. Chia et al. [12] proposed an image channel system to distill appropriate reference pictures from the gathered internet pictures. It requires the user to give a semantic content mark to look for a good reference picture on the Internet and human-division signals for the frontal area objects. Rather than the past colorization techniques, the proposed strategy is completely programmed as it uses an enormous arrangement of reference pictures from various classes with different objects.

Example-based colorization via *deep learning*—Cheng et al. [13] introduced a technique to colorize pictures dependent on deep learning. Deshpande et al. [14] explored a LEARCH system to prepare a quadratic target work in the chromaticity guides and perform colorization by limiting this goal work. Larsson et al. [15] utilized the hypercolumns of the VGG-16 engineering for getting highlights for accomplishing the shading move. Zhang et al. [3] introduced an automatic convolutional neural network (CNN)-based methodology for colorization while representing the issue as an order task. They utilized the class-rebalancing method at preparing time to expand the decent variety of hues in the outcome. Iizuka et al. [16] proposed a deep learning approach that included a combination layer that empowered rich consolidating of nearby data with global priors processed utilizing the whole image. Liu et al. [17] proposed a colorization model for single satellite imagery using multi-task deep neural networks.

The other conventional methods include the model proposed by Hussein and Yang [18] which utilized the edge preserving smoothing filter for colorization. Ju et al. [19] suggested a color fringe removal in narrow color regions of digital images. Dong et al. [20] proposed a self-supervised colorization model using cycle CNN. Kong et al. [21] suggested semantic segmentation-based adversarial edge-aware image colorization model. Nguyen-Quynh et al. [22] proposed an image colorization model using global scene-context style and pixel-wise semantic segmentation.

3 Proposed Model

The proposed model involves three phases. In the first phase, a pre-trained colorization model, as suggested by Zhang et al. [2], is deployed to color a grayscale image. Then, the colors spilled along the edges are identified in the second phase by choosing an optimal number of color clusters. Finally, in the third phase, user-guided image colorization, as suggested by Zhang et al. [3], is done by considering the centroids generated in the second phase to improve the aesthetics of the image. Figure 3 shows the workflow of the proposed model.

3.1 Colorful image colorization

Zhang et al. [2] proposed a wholly programmed approach that produces dynamic, what is more, practical colorizations. They grasped the fundamental vulnerability of the issue by acting it as a classification errand and used class-rebalancing at training time to build the assorted variety

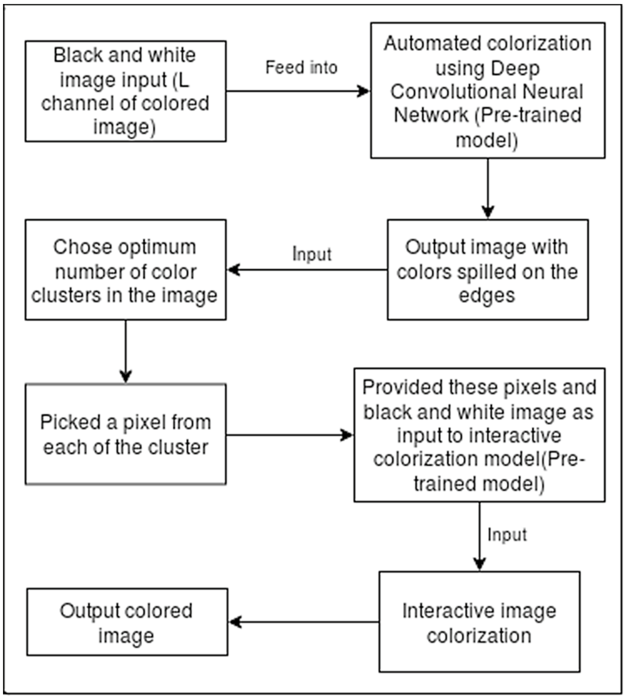


Fig. 3 The workflow of the proposed model

of hues in the outcome. The framework is actualized as a feed-forward go in a CNN, as shown in Fig. 4, at test time, and is trained on over a million shading pictures using ImageNet. Given the channel L, $\mathbf{X} = \mathbb{R}^{H\times W\times 1}$, the model needs to predict the corresponding a and b color channels, $\widehat{\mathbf{Y}} = \mathbb{R}^{H\times W\times 2}$, of the image in the CIE Lab colorspace, where H and W represent the height and width of the image, respectively. It can be achieved by minimizing the Euclidean loss between the predicted $\hat{\mathbf{Y}}$ and ground truth $\mathbf{Y}$ colors computed as

$$L_2\left(\hat{\mathbf{Y}}, \mathbf{Y}\right) = \frac{1}{2}\sum_{h,w}\left\|\mathbf{Y}_{h,w} - \hat{\mathbf{Y}}_{h,w}\right\|_2^2 \tag{1}$$

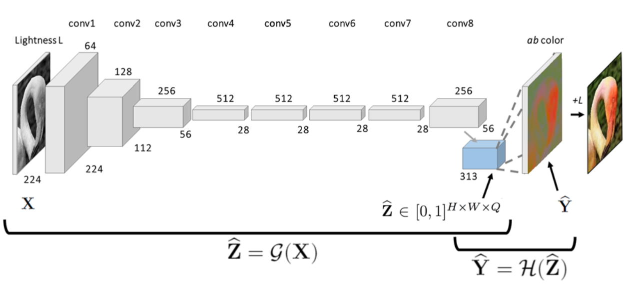


Fig. 4 CNN architecture of colorful image colorization [2]

3.2 Selection of optimal color clusters

The output of phase one produces spatial inconsistencies, as shown in Fig. 1. The colors are spilled along the edges. This can be removed by recoloring each object of the image with the appropriate color, which leads to another challenging problem upon the selection of colors per object.

An image contains objects of different colors; selecting one patch for each color is sufficient to be given as input to phase three of the proposed model. To determine the number of colors in the image, i.e., the number of patches to be taken, clustering was performed on the image based on the RGB values of the pixels. K-means clustering is used to perform the clustering [23]. Since different images can have a different number of colors, the number of clusters is different for different images. Thus, for every image, we applied K-means clustering for different values of k ranging from c_1 to c_2, where c_1 and c_2 are determined experimentally. In our case, $c_1 = 5$ and $c_2 = 21$. Then, Davies–Bouldin index (DBI) [24] is used to determine the goodness of clusters for different k. DBI is employed (chosen for evaluating clustering goodness) because of its comparatively more straightforward calculation and because it is computed based only on the quantities and features inherent to the dataset. The graph is shown in Fig. 5 establishes that neither a very low nor very high value of k produces optimum clustering. The value of k for which DBI is optimum (minimum) is found at $k = 6$ for the image shown in Fig. 5. Similarly, the results are found for other images. Thus, for every image, we ran K-means clustering for a different number of clusters and chose the one with a minimum value of DBI. The DBI_k for k number of clusters is given as

$$\mathrm{DBI}_k = \frac{1}{k}\sum_{i=1}^{k}\left(\max_{j=1\ldots k, i\neq j}\left(\frac{\mathrm{intra}_i + \mathrm{intra}_j}{\mathrm{inter}_{ij}}\right)\right) \tag{2}$$

where

$$\mathrm{intra}_i = \frac{1}{n_i}\sum_{p\in\mathbf{P}_i}\mathrm{distance}\left(\mathbf{I}(p), \overline{\mathbf{x}}_i\right) \tag{3}$$

$$\mathrm{inter}_{ij} = \mathrm{distance}\left(\overline{\mathbf{x}}_i, \overline{\mathbf{x}}_j\right) \tag{4}$$

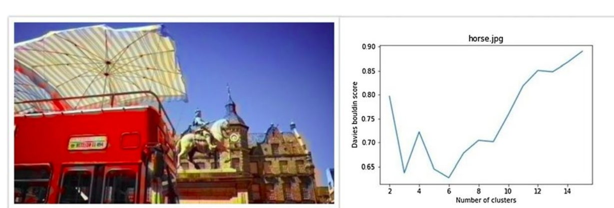


Fig. 5 Evaluation of optimal no. of clusters for a sample image

$$\boldsymbol{\mu}_i = \frac{1}{n_i} \sum_{x \in \mathbf{X}_i} \mathbf{I}(x) \tag{5}$$

where x is a pixel in the input RGB color image $\mathbf{I}$, $\mathbf{X}_i$ is the set of pixels belonging to the ith cluster, n_i is the number of pixels belonging to the ith cluster, and $\overline{\mathbf{x}}_i$ is the centroid of the ith cluster. $distance(\mathbf{a}, \mathbf{b})$ measures distance between vector $\mathbf{a}$ and vector $\mathbf{b}$ using the L2 norm. intra_i evaluates the compactness (intracluster distance) of the ith cluster, and inter_{ij} is the inter-cluster distance between cluster i and cluster j. The optimal value of the number of clusters k^* can be achieved by selecting the clustering results corresponding to the minimum value of DBI_k and is given as

$$k^* = \arg\min_k \text{DBI}_k \tag{6}$$

Now that we have the clusters, we have to choose a patch from every cluster. The steps involved in choosing the patches are as follows:

a. For cluster i, the point nearest (in terms of Euclidean distance) to its centroid is found. Let this point be P_i. Since the centroid of a cluster may or may not be in the population. So, this step was performed to find a representative of the population from the population itself.
b. Then for this cluster, the points nearest to P_i are taken. Specifically, the points at a distance of [0, 1) from P_i are taken. Let the set of these points be S_i.
c. The points in S_i are RGB values of pixels. For every point in S_i, we find the point corresponding to that RGB color value. Let this set of points be C_i.
d. Then C_i is sorted in ascending order, first by x-coordinate and then by y-coordinate.
e. Then the median of sorted set C_i is taken, and a 3×3 patch with this point as a center is selected as input patch.

The steps a–e are repeated for every cluster in the image, and finally, all the generated patches are given as input to phase three along with the black and white image to generate the final output. An overview of it can be seen in Fig. 6.

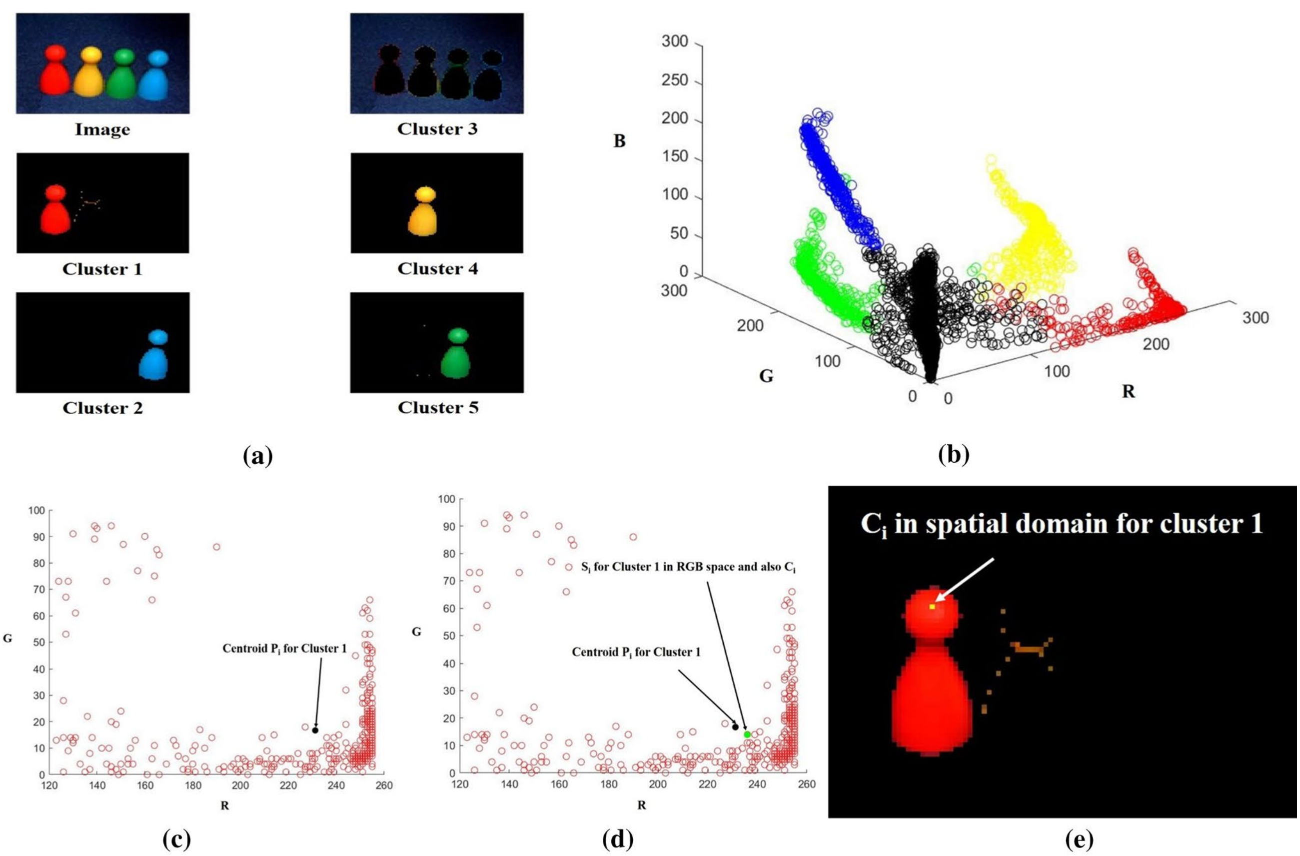


Fig. 6 **a** DBI-based clustering results in spatial domain, **b** DBI-based clustering results for every pixel in RGB color space, **c** identification of P_i for cluster 1 (red colored object) in RGB color space, **d** identification of C_i for cluster 1 in RGB color space, and **e** identification of C_i for cluster 1 in spatial domain

3.3 User-guided image colorization

Zhang et al. [3] proposed a user-guided colorization approach to combine the best of both worlds between deep automatic approaches and user-guided approaches. They proposed a system that has information about raw image priors learned from data, free to predict color even in the absence of explicit user inputs, as well as the ability to integrate low-level texture or high-level semantic cues to propagate user edits efficiently. Training the automatic method can be treated as a typical supervised learning problem. Given grayscale color pairs $(\mathbf{X}, \mathbf{Y})$, one could look at the color prediction $F(\mathbf{X})$, ground truth color $\mathbf{Y}$, and measure the distance, or loss, between them. They adjusted the parameters of network F to minimize this loss across the training set.

A similar approach with a user-guided network was utilized. However, there is a significant problem here. User points are challenging to obtain. While one can simply download millions of images and break them up into their grayscale and color components to make $\mathbf{X}$ and $\mathbf{Y}$, user interaction data is not so readily available. Collecting millions of interactions is likely to be expensive. And even more nefarious, the user behavior would be dependent on the system itself. And of course, the system behavior is dependent on the user training data, so it is like a chicken and egg problem. This can be solved by randomly simulating users. They have taken the ground truth $\mathbf{Y}$, and randomly reveal points, and treat this as the simulated user points $\mathbf{U}$. These randomly revealed points are viewed as "hints". The network gets a small peak at ground truth $\mathbf{Y}$ and tries to guess the rest of it. The CNN architecture utilized is shown in Fig. 7, and the objective is to select an appropriate color prediction that minimized the loss between the ground truth and the color prediction based on user points.

$$F^* = \arg\min_F l(F(\mathbf{X}, \mathbf{U}), \mathbf{Y}) \tag{7}$$

The performance of this phase highly depends upon the parameter $\mathbf{U}$. So, in this proposed model, the result of phase two is fed as parameter $\mathbf{U}$.

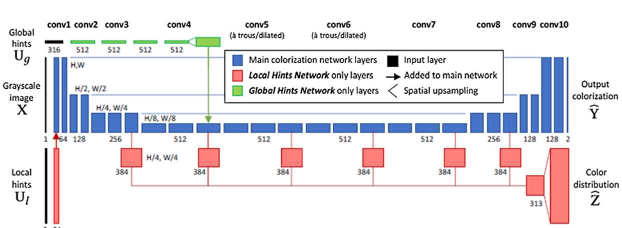


Fig. 7 CNN architecture of user-guided image colorization [3]

4 Experimental setup and results

In this section, the evaluation of the model on test datasets is outlined. All analyses are performed utilizing NVidia GeForce GTX 1060 GPU, Intel Core i5 CPU with 8 GB RAM, and Python libraries.

4.1 Dataset used

The trials are led utilizing the ImageNet dataset [25], with different object categories, as shown in Table 1. The testing was done over 10 k images randomly containing different objects/scenes from the ImageNet dataset.

4.2 Qualitative evaluation

The qualitative analysis of the proposed model compared to the benchmark models suggested by Zhang et al. [2, 3]

Table 1 ImageNet dataset [25]

High-level category	# synset (subcategories)	Avg # images per syncet	Total # images
Amphibian	94	591	56 K
Animal	3822	78	5415
Appliance	51	1164	59 K
Bird	856	949	812 K
Covering	946	819	774 K
Device	2385	675	1610 K
Fabric	262	690	181 K
Fish	566	494	280 K
Flower	462	735	339 K
Food	1495	670	1001 K
Fruit	309	607	188 K
Fungus	303	453	137 K
Furniture	187	1043	195 K
Geological formation	151	838	127 K
Invertebrate	728	573	417 K
Mammal	1138	821	934 K
Musical instrument	157	891	140 K
Plant	1666	600	999 K
Reptile	268	707	190 K
Sport	166	1207	200 K
Structure	1239	763	946 K
Tool	316	551	174 K
Tree	993	568	564 K
Utensil	86	912	78 K
Vegetable	176	764	135 K
Vehicle	481	778	374 K
Person	2035	468	952 K

Fig. 8 Comparison based upon qualitative analysis

can be seen in Fig. 8. Images containing different objects/scenes were chosen to show the analysis. It can be seen that the spilling out of colors, as seen in the model suggested by Zhang et al. [2] has been removed. The colors chosen by the proposed model are better than the default colors selected by Zhang et al. [3] compared to the ground truth (colored image). Our model generates more plausible and spatially consistent images, which is evident from comparing the results in Fig. 8.

4.3 Quantitative evaluation

We calculated the mean square error (MSE) and structural similarity (SSIM) between the ground truth and colorized output on 10 k images in the ImageNet validation set [22] for quantitative evaluation, given as

$$\mathrm{MSE}\left(\hat{\mathbf{I}}, \mathbf{I}\right) = \frac{1}{H \times W} \sum_{H,W} \left\| \mathbf{I}_{H,W} - \hat{\mathbf{I}}_{H,W} \right\|_2^2 \quad (8)$$

$$\mathrm{SSIM}\left(\hat{\mathbf{I}}, \mathbf{I}\right) = \frac{\left(2\mu_{\hat{\mathbf{I}}}\mu_{\mathbf{I}} + C_1\right)\left(2\sigma_{\mathbf{I}\hat{\mathbf{I}}} + C_2\right)}{\left(\mu_{\hat{\mathbf{I}}}^2 + \mu_{\mathbf{I}}^2 + C_1\right)\left(\sigma_{\hat{\mathbf{I}}}^2 + \sigma_{\mathbf{I}}^2 + C_2\right)} \quad (9)$$

where $\mathbf{I}$ and $\hat{\mathbf{I}}$ are ground truth and predicted image, respectively, H, W being their height and width. $\mu_{\mathbf{I}}$ and ${\sigma_{\mathbf{I}}}^2$ are the mean and variance of ground truth, respectively, while $\mu_{\hat{\mathbf{I}}}$ and ${\sigma_{\hat{\mathbf{I}}}}^2$ are the mean and variance of predicted image, respectively. $\sigma_{\mathbf{I}\hat{\mathbf{I}}}$ is the covariance between ground truth and predicted image, $C_1 = C_2 = 0$.

The MSE and SSIM of the proposed model and other related models are shown in Table 2. The best result is marked in bold. The following can be inferred from Table 2.

- The proposed model outperforms the state-of-the-art models in terms of MSE with an improvement of 51.2%, 30.6%, 21.2%, 14.3%, 19.1%, and 12.3% over the models suggested by Levin et al. [4], Luan et al. [8], Gupta et al. [10], Chia et al. [12], Zhang et al. [2], and Zhang et al. [3], respectively.
- The proposed model outperforms the state-of-the-art models in terms of SSIM with an improvement of 27.3%, 18.4%, 17%, 11.9%, 13.5%, and 5.1% over the models suggested by Levin et al. [4], Luan et al. [8], Gupta et al. [10], Chia et al. [12], Zhang et al. [2], and Zhang et al. [3], respectively.

5 Conclusion

In this paper, a fully automated system is proposed for grayscale image colorization to remove spatial inconsistencies present in the image after colorization. To achieve it, first, the grayscale image is colored using a standard CNN-based model. Then, the colors of different objects present in the image are identified, and the image is recolored based upon the colors identified. This helps in the removal of any color spills and spatial inconsistencies. The proposed model was compared with the state of the art over ImageNet dataset in terms of mean square error and structural similarity. From the experiment, it is evident that the proposed model has outperformed the related works.

Table 2 Quantitative evaluation of proposed model and related work

Models	MSE	SSIM
Levin et al. [4]	412.4	0.6615
Luan et al. [8]	356.4	0.7430
Gupta et al. [10]	330.6	0.7558
Chia et al. [12]	311.7	0.8016
Zhang et al. [2]	324.8	0.7876
Zhang et al. [3]	306.4	0.8642
Proposed model	**272.8**	**0.9103**

The best results, like lowest value of MSE and highest value of SSIM, are marked in bold

References

1. Vieira, L.F.M., Vilela, R.D., Nascimento, E.R.D., Fernandes, F.A., Carceroni, R.L. and Araújo, A.D.A.: Automatically choosing source color images for coloring grayscale images. In: 16th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2003), pp. 151–158. IEEE (2003)
2. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: European Conference on Computer Vision, Amsterdam, Netherlands, pp. 649–666 (2016)
3. Zhang, R., Zhu, Y., Isola, P., Geng, X., Lin, A. S., Yu, T, Efros, A.A.: Real-Time User-Guided Image Colorization with Learned Deep Priors (2017). https://arxiv.org/abs/1705.02999v1
4. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. ACM SIGGRAPH 2004 Papers. Los Angeles, USA, vol. 14(1), pp. 689–694 (2004)
5. Huang, Y.C., Tung, Y.S., Chen, J.C. et al.: An adaptive edge detection based colorization algorithm and its applications. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, Singapore, vol. 14(1), pp. 351–354 (2005)
6. Yatziv, L., Sapiro, G.: Fast image and video colorization using chrominance blending. IEEE Trans. Img. Proc. **15**(5), 1120–1129 (2006)
7. Qu, Y., Wong, T.-T., Heng, P.-A.: Natural image colorization. ACM SIGGRAPH 2006 Papers, SIGGRAPH 06, Boston, USA, vol. 3, pp. 1214–1220 (2006)
8. Luan, Q., Wen, F., Cohen-Or, D., et al.: Natural image colorization. In: Proceedings of the 18th Eurographics Conf. on Rendering Techniques, EGSR 07, Grenoble, France, vol. 42(3), pp. 309–320 (2007)
9. Charpiat, G., Hofmann, M., Scholkopf, B.: Automatic image colorization via multimodal predictions. In: European Conference on Computer Vision ECCV, pp. 126–139 (2008)
10. Gupta, R.K., Chia, A.Y.-S., Rajan, D., et al.: Image colorization using similar images. ACM Int. Conf. Multimedia **1**, 369–378 (2012)
11. Liu, X., Wan, L., Qu, Y., et al.: Intrinsic colorization. Trans. Graph. **27**, 152 (2008)
12. Chia, A.Y.-S., Zhuo, S., Gupta, R.K., et al.: Semantic colorization with internet images. ACM Trans. Graph. **30**(156), 1–8 (2011)
13. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: International Conference on Computer Vision (ICCV), Las Condes, Chile, vol. 1 (2015)
14. Deshpande, A., Rock, J., Forsyth, D.: Learning large-scale automatic image colorization. In: International Conference on Computer Vision (ICCV), Las Condes, Chile, vol. 1 (2015)
15. Larsson, G., Gustav, M.M., Shakhnarovich, G.: Learning representations for automatic colorization. In: European Conference on Computer Vision, Amsterdam, Netherlands, pp. 577–593 (2016)
16. Iizuka, S., Edgar, S.S., Ishikawa, H.: Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Trans. Graph. (TOG) **35**(4), 110 (2016)
17. Liu, H., Fu, Z., Han, J., Shao, L., Liu, H.: Single satellite imagery simultaneous super-resolution and colorization using multi-task deep neural networks. J. Vis. Commun. Image Represent. **53**, 20–30 (2018)
18. Hussein, A.A., Yang, X.: Colorization using edge-preserving smoothing filter. SIViP **8**(8), 1681–1689 (2014)
19. Ju, H.J., Lee, D.K., Park, R.H.: Color fringe removal in narrow color regions of digital images. SIViP **8**(8), 1651–1662 (2014)
20. Dong, X., Liu, C., Li, W., Hu, X., Wang, X., Wanga, Y.: Self-supervised colorization towards monochrome-color camera systems using cycle CNN. IEEE Trans. Image Process. **30**, 6609–6622 (2021)
21. Kong, G., Tian, H., Duan, X., Long, H.: Adversarial edge-aware image colorization with semantic segmentation. IEEE Access **9**, 28194–28203 (2021)
22. Nguyen-Quynh, T.T., Kim, S.H., Do, N.T.: Image colorization using the global scene-context style and pixel-wise semantic segmentation. IEEE Access **8**, 214098–214114 (2020)
23. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics (2007).
24. Davies, D. L., Bouldin, D. W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-1**(2): 224–227 (1979)
25. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint https://arxiv.org/abs/1409.1556