Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Colorization of Digital Images: An Automatic and Efficient Approach through Deep learning

S. J. Sugumar

Professor, Department of Electronics and Communication Engineering, MVJ College of Engineering, Bengaluru, India

E-mail: sugumar.sj@mvjce.edu.in

Abstract

Colorization is not a guaranteed, but a feasible mapping between intensity and chrominance values. This paper presents a colorization system that draws inspiration from recent developments in deep learning and makes use of both locally and globally relevant data. One such property is the rarity of each color category on the quantized plane. The denoising model contains hybrid approach with cluster normalization through U-Net deep learning construction of framework. These are built on the basic U-Net design for segmentation. To eliminate gaussian noise in digital images, this article has developed and tested a generic deep learning denoising model. PSNR and MSE are used as performance measures for comparison purposes.

Keywords: U-Net, noise removal, colorization, de-noising, deep learning, convolutional neural network

1. Introduction

When it comes to computer vision and the study of pictures, one of the most difficult problems is real-time image categorization. In the field of color image processing, several color models are now in use. Color models like the HSV (Hue Saturation Value) scheme are preferred over the RGB (Red, Green, Blue) color space that naturally results from color display hardware due to their supposedly more user-friendly nature. Using the RGB color paradigm, pictures are broken down into their individual red, green, and blue components. However, from the perspective of how humans see colors, the RGB format is not ideal for representing visual content. Color also isn't made up just of the main hues red, green, and

DOI: https://doi.org/10.36548/jiip.2022.3.006
Received: 27.07.2022, received in revised form: 20.08.2022, accepted: 06.09.2022, published: 16.09.2022

blue [1-5]. The human visual system uses the contrast and color of an item to determine what it is. Saturation and color temperature are used to characterize the latter. A person's perception of how bright something is, may vary greatly. The achromatic concept of intensity is realized there. Hue, an aspect of color, stands for "main color." The saturation of a color describes how much white light has watered down its intensity. Because of its similarities to the human visual system, the HSV model is based on it. The HSV model separates the luminous component (brightness) from the informational role of color (hue and saturation). As a result, the HSV model of describing colors is more intuitive for the human visual system than the RGB model [6, 7]. Figure 1 shows some example of colorization of digital images.

The process of colorizing is performed on a black-and-white object. The method of colorization is one example of post-production used on digital photos. Regrettably, this method has a long history of being derided as time-consuming and boring. Despite recent advances in automation, achieving a desirable colorization outcome sometimes still requires extensive human intervention [8-10].

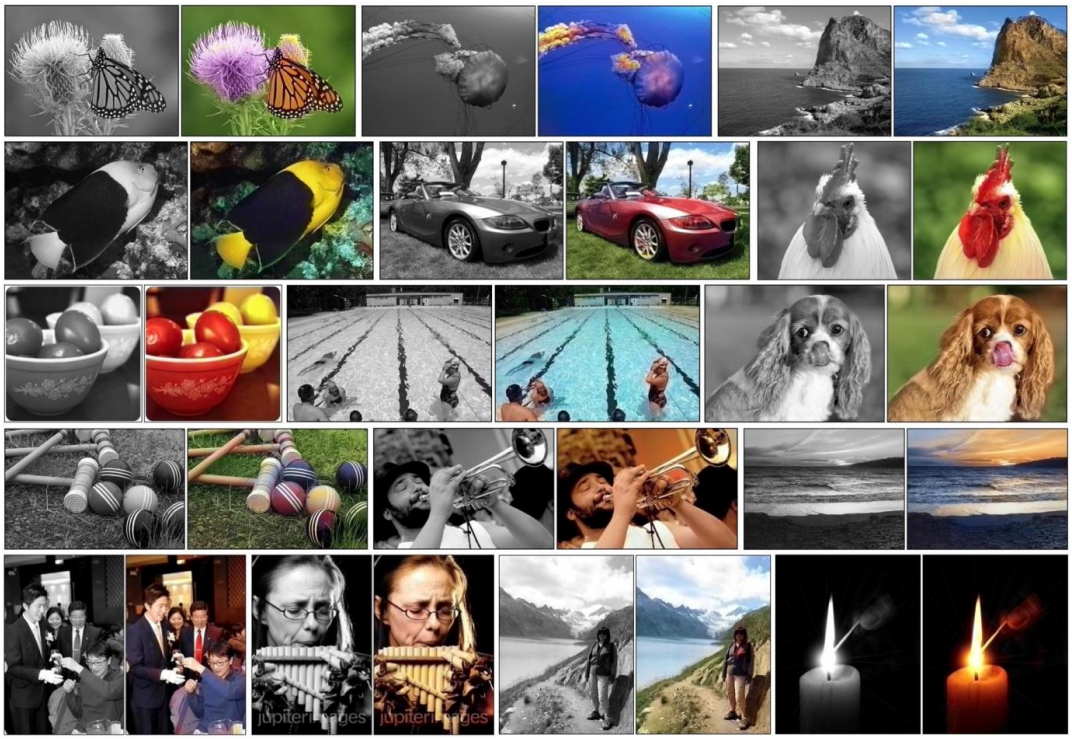

Figure 1. Example of Colorization of Digital Images [25]

During the processes of capturing, compressing, storing, and transmitting digital photos, unwanted noise is introduced. Smoothing and thresholding in an adapted domain are often used in computer vision research to restore the original, unaltered picture [1]. Taking a noisy picture as input, a denoising algorithm will produce an image with much less noise [2].

Denoising images serves a far larger function in the field of image processing than just making them seem better. For more complex computer vision tasks like classification, segmentation, and object identification, "denoising" is a solution component used to boost performance.

The two core techniques for picture denoising are filtering and wavelet modification. The amalgamation is performed using a combination of Gaussian and bilateral filters, and thresholding is performed utilizing wavelets [11]. Filtering methods may be used to smooth down rough edges and diminish the presence of noise. However, edge information is lost during the smoothing process, and the addition of noise through filters alters the colors such that they no longer correspond with the original [12].

2. Literature Survey

In recent years, there have been a lot of ground-breaking study into how to best colorize digital images.

Colorization methods described by Levin et al. [13] are predicated on the idea that contiguous pixels in space and time with equal intensities should have similar hues. For this, they made use of a cost function that is quadratic in nature. By using this technique, an artist or user may annotate a picture with a few scribbles of color, and those hues will be applied uniformly throughout the whole scene or animation. The methodology presented in [14] makes advantage of automated colorization algorithms to provide vivid and realistic colorizations. Also, they treated it as a classification issue and employed class-rebalancing during training to make the output more colorful.

To automatically add color to black-and-white photos, Izuka et al. [15] offered a colorization technique that combines global priors with local image characteristics and categorization. To combine the advantages of both categorizations, their deep network made use of a fusion layer to complete the joining process. To get quality free of artifacts, Cheng et al. [16] suggested a post-processing phase based on combined bilateral filtering. To take into account the worldwide picture data, they also created an adaptive image clustering method.

According to Deshpande et al. [17], the chromaticity maps used a quadratic objective function, and automated colorization is a regression in a continuous ab color plane. It was possible to regulate the spatial inaccuracy in the colorization of the picture using the objective

function, which allows for correlations on large spatial scales. When coloring an image, this goal function was minimized.

The two core techniques for picture denoising are filtering and wavelet modification. Denoising digital photos, however, has lately seen a new technique emerge: machine learning. A trained autoencoder may produce new pictures that are almost identical to the ones it is fed as input. This design minimizes the number of dimensions in the incoming picture data by learning identity mapping. Based on the Autoencoder architecture for learning, the Denoising Autoencoder is trained to produce clean anticipated pictures from noisy input. Despite minor, unimportant shifts in the topics under examination, the Denoising Autoencoder nonetheless strives to achieve a structured input distribution.

DenseNet, suggested by Huang et al. [18], maximizes information flow across layers by connecting them through a series of channels formed from the outputs of all levels. U-Net with Gn nodes include a DenseNet component to provide a direct comparison of Residual Net and DenseNet with regard to their respective shortcut connection properties. DenseNet-style, this network builds upon itself by chaining together the results of each successive layer.

2.1 Motivation of the Research

In light of these researches, it is decided to look into the possibility of using a reconstructed U-Net architecture to learn universal denoising. As part of this research, three distinct denoising models are built using the U-Net architecture with clustering utilising normalization techniques. This article details their arrangements and how they enhance the features work.

3. Methodologies

Exogenous excitatory (extrinsic) and inhibitory (intrinsic) impulses, as well as inter-neuronal (intrinsic) communication, all influence the behavior of the n-neuron neural network. In both the classical and modern models, neuronal firing results in the transmission of excitatory and inhibitory signals to neighboring neurons and the outside environment, respectively.

3.1 Scribble-based colorization

Here, the user colors on the monochrome picture by scribbling on it with various colors, and the computer then colors the image according to the user's choices. As a result,

areas of equal intensity may be distinguished from one another by their shared hue. The time and effort required by the user to produce such nearly precise color scribbles makes this approach impractical.

3.2 Coloring by Examples

In order to learn the tonal qualities of an input picture, example-based approaches look to a reference image. The authors of [19] used segmentation in order to figure out the color of the portion of the picture, each pixel should take its hue cues from the example picture. This is accomplished mechanically by the use of a strong supervised classification strategy that examines the pixel-by-pixel feature space included in the sample picture. Example-based colorizations may be further classified by the origin of the reference image.

- The user must provide a good reference picture for the colorization to work with option. It uses the intensity and surrounding characteristics of the pixels to locate a matching one in the reference picture, and then applies that pixel's color to the target pixel.
- The user is spared the trouble of choosing an appropriate picture via this method. Colorization utilizing examples from the web, makes advantage of the vast amount of image data available on the Internet. An intrinsic picture is created by Liu et al., by combining a set of Internet reference photos that are visually comparable to one another.

3.3 U-Net

It is not the U-original Net's structure, hence the input and output picture sizes are different. The model has been recreated by adjusting the parameters of the convolution function in order to compare the quality of denoising for images of the same size.

3.4 Denoising using a U-Net Framework

As the deep denoising model, this research relies on three variants of the deep encoder-decoder model U-Net. Skip links between narrowing and widening pathways are also included in the U-Net. Raw grayscale values, DAISY features [20], and High-level semantic features are all retrieved from each pixel in the training pictures. Next, a deep neural network is trained using the combined features.

3.5 U-Net with group normalization

Including batch normalization during training improves optimization. However, a big enough batch size is required for batch normalization. Consequently, Batch Normalization has been opted while developing the U-Net and experimented with denoising after training. Unfortunately, the testing process revealed that the short batch size chosen prevented from achieving satisfactory denoising results. This is why group normalization [21] has been used and implemented in the U-Net at every level.

Instead of normalizing all the channels in a feature map at once, as is done in batch normalization, groups of channels are set and normalized independently. For this normalization to achieve the same outcomes as batch normalization, a small batch size is sufficient. Figure 2 & 3 shows the block diagrams of overall proposed architecture.

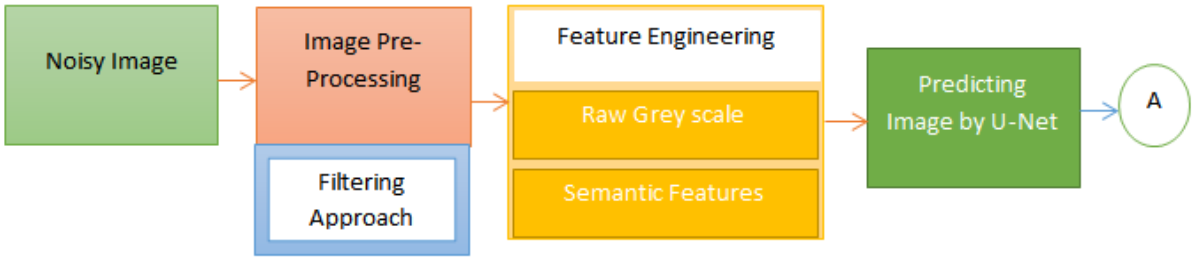

Figure 2. Overall proposed architecture – Phase 1

3.6 Applying Denoising Methods in Training

The deep-learning framework is used to implement denoising learning. Processors found in graphics cards are used to train the network. Each model is trained to perform blind denoising using 90,067 training patches over the course of 30 epochs using a batch size of 5, the Adam optimization function, and a standard deviation of 0.05.

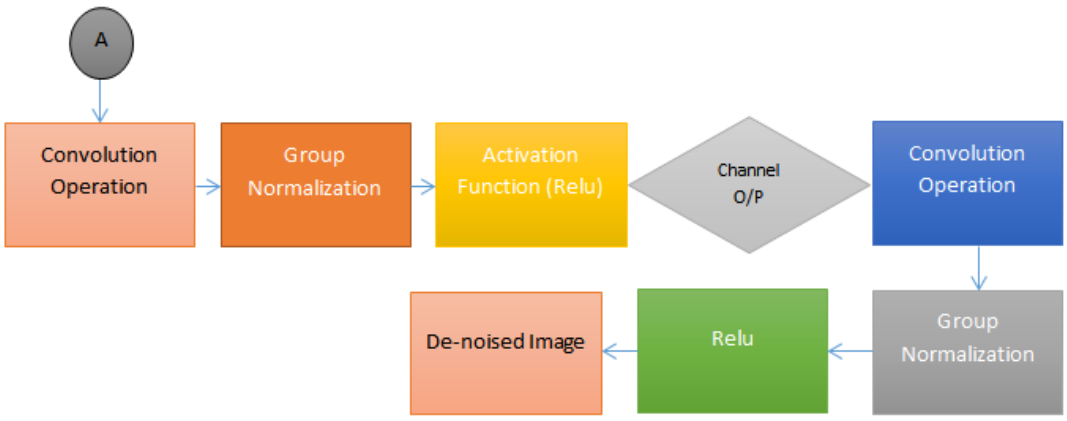

Figure 3. Overall proposed architecture – Phase 2

4. Results and Discussion

4.1 Dataset

The ADE20K image dataset [22] is used to test and train the proposed denoising algorithms on noisy pictures with varying levels of noise. Semantic scene segmentation in photographs is the focus of the ADE20K dataset. The settings range is from inside to outdoors to urban environments. Compared to COCO and ImageNet [23], this dataset has a much larger number of object types. Twenty-one thousand and ten photos from the training set and three thousand and fifty-two from the testing set are employed in this experiment. Figures 4a and 4b show the accuracy and loss charts of the proposed work.

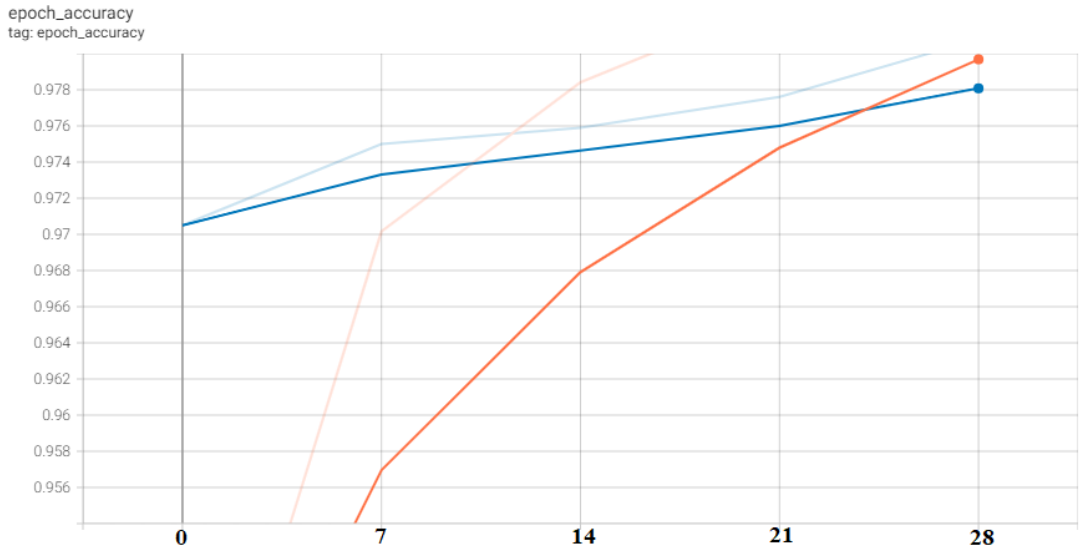

Figure 4a. Accuracy chart of the proposed technique

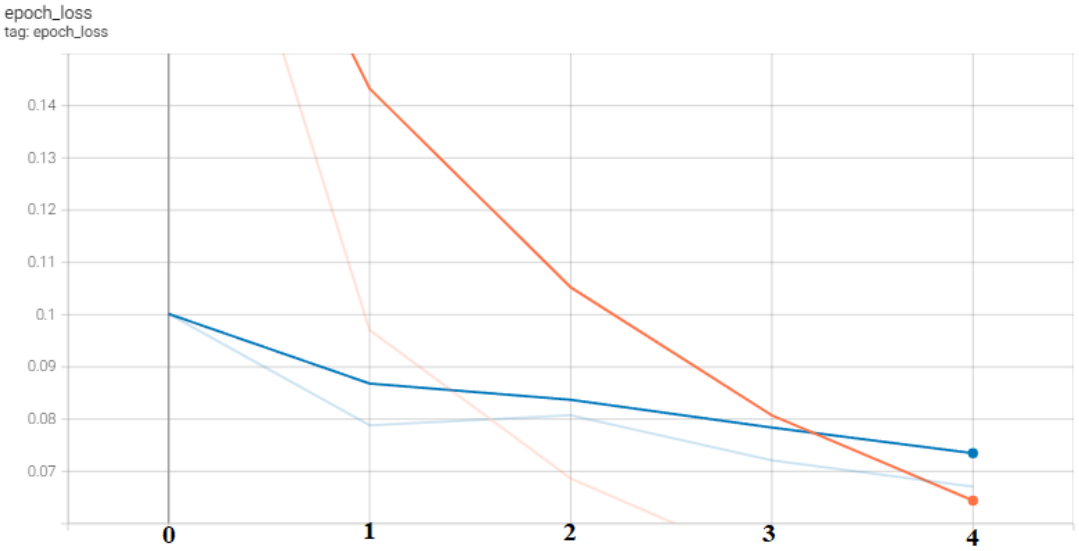

Figure 4b. Loss Chart of the proposed technique

4.2 Image Pre-Processing

For preparation work like summation, the OpenCV computer vision library is employed. Modifying the picture by adding or subtracting noise, and by constructing patches out of the whole thing is performed. After training each model for denoising, test dataset photos are used to assess the quality of the denoised results. PSNR is used to measure the quality of the images and SSIM to measure the degree to which the predicted picture resembles the original.

4.3 Generating Patches from the Image

The cropping of input images through patch generation process and obtaining the target image down to 256 by 256 patches allows to work with images of varying dimensions (the image before processing). 90k patches have been obtained from 20k photos by repeatedly cropping the images in the training set. The expected picture is constructed by piecing together individual patches generated by a deep denoising algorithm.

One major drawback of scribble-based methods is that their performance depends on the extent of human involvement. An acceptable color picture must be provided as one of the inputs to the algorithm in order for the example-based colorization to be successful. In the proposed method, a training dataset is produced, and a neural network is trained to predict the values, similar to what was done in [24]. In contrast to it, the semantic information is extracted entirely mechanically. Figure 5 shows the overall efficiency of the system from the obtained table 1.

Table 1. Comparison results between methodologies

Methods	Accuracy	Precision	Sensitivity	Loss	PSNR	MSE
Scribble Based Colorization	81%	80.2%	85.2%	0.10	26.9	0.821
Conventional CNN	82%	79.6%	84%	0.11	28.3	0.828
Proposed U-net Architecture	94.3%	96.2%	93%	0.03	31.2	0.621

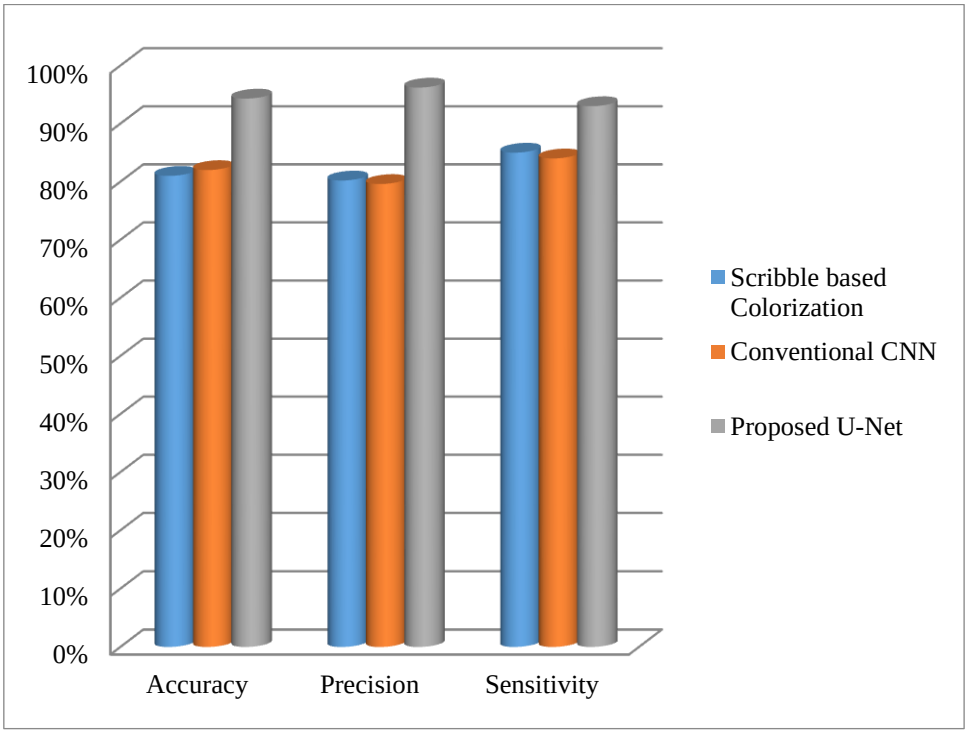

Figure 5. Overall Efficiency of the proposed system

4.4 Denoising Performance of the U-Net based model compared to the default models

The denoising capabilities of this model are compared to those of several industry standards. The PSNR values for different kinds of noise are shown in separate vertical portions of the table. Maximum signal strength divided by the strongest noise that degrades the signal's representation is known as the Peak Signal-to-Noise Ratio (PSNR). In most cases, a greater PSNR indicates a more accurate reconstruction.

5. Conclusion

The proposed approach is successful against high levels of visual noise but fails miserably when it comes to identifying lower levels of noise or noise that is not apparent at all. Denoising learning using datasets with undetectable soft noise is required to get beyond this flaw. Reducing noise has far-reaching effects beyond improving visual aesthetics. High-quality pictures are crucial to the success of image processing applications including medical imaging, autonomous driving, and robotics research. As a result, picture denoising plays a crucial part in contemporary image processing infrastructure. Denoising the photos using the created model allows for more precise object identification and photo detection by reducing characteristics affected by heavy noise. Using more datasets and transfer learning, how to

further refine the fundamental denoising model can be studied. As a consequence of this, it will be more precise and practical.

References

[1] Y. Tian, P. Luo, X. Wang, and X. Tang. Deep learning strong parts for pedestrian detection. Proceedings of the IEEE International Conference on Computer Vision, 1904-1912, 2015.

[2] D. Varga, T. Szir_anyi, A. Kiss, L. Sp_or_as, and L. Havasi. A multi-view pedestrian tracking method in an uncalibrated camera network. Proceedings of the IEEE International Conference on Computer Vision Workshops, 37-44, 2015.

[3] D. Ciresan and U. Meier. Multi-column deep neural networks for offline handwritten Chinese character classification. Proceedings of the International Joint Conference on Neural Networks, 1-6, 2015.

[4] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (TOG), 35(4): 110, 2016.

[5] Tokui, S.; Okuta, R.; Akiba, T.; Niitani, Y.; Ogawa, T.; Saito, S.; Suzuki, S.; Uenishi, K.; Vogel, B.; Vincent, H.Y. Chainer: A Deep Learning Framework for Accelerating the Research Cycle. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 3–7 August 2019.

[6] Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.

[7] Li, C.-N.; Shao, Y.-H.; Deng, N.-Y. Robust L1-norm two-dimensional linear discriminant analysis. Neural Netw. 2015, 65, 92–104.

[8] Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A. Scene Parsing through ADE20K Dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

[9] Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Computer Vision—ECCV 2016; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2014; pp. 740–755.

[10] López-Tapia, S.; Lucas, A.; Molina, R.; Katsaggelos, A.K. A single video super-resolution GAN for multiple down sampling operators based on pseudo-inverse image formation models. Digit. Signal Process. 2020, 104, 102801.

[11] Bell-Kligler, S.; Shocher, A.; Irani, M. Blind super-resolution kernel estimation using an internal-GAN. Adv. Neural Inf. Process. Syst. 2019, 1, 284–293.

[12] Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

[13] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. ACM Transactions on Graphics, 23(3): 689{694, 2004.

[14] Deshpande, Aditya, Jason Rock, and David Forsyth. ”Learning large-scale automatic image colorization.” Proceedings of the IEEE International Conference on Computer Vision. 2015.

[15] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classi_cation. ACM Transactions on Graphics (TOG), 35(4): 110, 2016.

[16] Z. Cheng, Q. Yang, and B. Sheng. Deep colorization. Proceedings of the IEEE International Conference on Computer Vision, 415-423, 2015.

[17] A. Deshpande, J. Rock, and D. Forsyth. Learning Large-Scale Automatic Image Colorization. Proceedings of the IEEE International Conference on Computer Vision, 567-575, 2015.

[18] Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

[19] Li, Y.; Zhang, B.; Florent, R. Understanding neural-network denoisers through an activation function perspective. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.

[20] Burger, H.C.; Schuler, C.J.; Harmeling, S. Image denoising: Can plain Neural Networks compete with BM3D? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, Rhode Island, 16–21 June 2012.

[21] Abubakar, A.; Zhao, X.; Li, S.; Takruri, M.; Bastaki, E.; Bermak, A. A Block-Matching and 3-D Filtering Algorithm for Gaussian Noise in DoFP Polarization Images. IEEE Sens. J. 2018, 18, 7429–7435.

[22] Farabet, Clement, et al. ”Learning hierarchical features for scene labeling.” IEEE transactions on pattern analysis and machine intelligence 35.8 (2012): 1915-1929.

[23] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research 15 (2014), pp: 1929–1958.

[24] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, Caffe: Convolutional Architecture for Fast Feature Embedding, ACM International Conference on Multimedia 2014, pp: 675–678.

[25] Website: https://richzhang.github.io/colorization/

Author's biography

S. J. Sugumar works as a Professor in the Department of Electronics and Communication Engineering, MVJ College of Engineering, Bengaluru, India. His area of research includes embedded system, IoT, WSN, artificial intelligence, image processing.