Colourify: Image Colourization using AI techniques

Eeshita Deepta[1], Neha Juyal[2], Khushi Pareek[3], Bhawna Jha[4], Dr Sumita Gupta[5]

[1,2,3,4,5]Amity University, Noida
{edeepta1713@gmail.com, juyalneha2@gmail.com, khushipareek76@gmail.com, bhawnajha1414@gmail.com, sgupta4@amity.edu}

Abstract. Converting a grayscale image into a coloured one is known as image colourization. This is a challenging task that needs human assistance to produce high-quality photos. Our model, which is an improved version of the VGG16, is what we suggest in this study. The dataset used is a subset of the MIT Places Dataset available on Kaggle. The architecture is divided into 2 parts, an automatic encoder and an automatic decoder. The model will predict the values of 'a' and 'b' in the Lab colour space. Architecture of the proposed model has been presented. The results have been compared with the existing models.

Keywords: Image Colourization, VGG-16, Deep Learning, Grayscale Images, CNN.

1 Introduction

During the colourization process, the most likely colour for every pixel in a grayscale image is predicted. In order to do this, the machine must be able to comprehend the semantic meaning of the items and their context in the image, which can be a difficult challenge [1]. However, the availability of sizable datasets, like the ImageNet dataset, as well as improvements in deep learning methodologies have made it possible to develop strong models that can precisely colourize images.

Numerous computer vision tasks, including semantic segmentation, object detection, and image classification, have made extensive use of well-known CNN designs like VGG-16 [2]. Their distinctive deep architecture, which is made up of a large number of convolutional and pooling layers, allows them to learn and capture intricate characteristics from the input image [1, 2]. Additionally, they excelled in image classification tasks, earning top rankings in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

VGG-16 is employed in certain studies as feature extractors to increase the precision of their colourization models, while in other studies it has served as the model's foundation [1,2,3]. In our model, the encoder is VGG-16. Numerous alternative coloured images may correlate to a single grayscale image since this problem is multimodal. As a result, traditional models usually needed more human input than just a grayscale image [4].

The study will assess the colourized pictures' degree of accuracy. The findings of this research can be utilized to direct the choice of the best CNN architecture for picture colourization tasks and offer insights into the advantages and disadvantages of the VGG-16, our modified VGG-16, and several other models.

2 Literature Review

In [1] paper, the ability to automatically colourize grayscale images is demonstrated via a completely automatic colourization system on VGG-16. Images can be made more bright and frequently contain colours that make sense thanks to the colour rebalancing. In other words, it sometimes corrects saturated colour that is undesired in some photographs.

In [2], the semantic information and image properties of an image that was processed by the CNN architecture were fetched using the VGG-16 pre-trained model. The LAB picture layers' A and B colours were anticipated, while L stands for the intensity layer, which is unchanged. The suggested concept has a chance of being adopted by media industries all around the world.

In paper [3] presented a brand-new, entirely automatic colourization method built on the VGG-16 and a two-stage CNN. A two-stage CNN architecture was trained using the numerous discriminative, sematic information offered by the VGG-16 without the usage of pooling layers. By including data from a previous layer, the two-stage design gave us a richer representation.

Paper [4] GHIM10K and CalTech256, two cutting-edge datasets, were used to fine-tune the network (AlexNet, VGG16, and VGG19) weight parameters for the image classification task. Utilizing the three criteria recall, precision, and F-score, compare the effectiveness of these network topologies. Furthermore, we used support vector machines for image classification to examine the robustness of CNN features. With the mentioned CNN networks, we have contrasted SVM performance.

In paper [5] deep learning techniques based on VGG-19 are employed for colourization. In order to forecast the colour of grey photos, it is necessary to make use of the latent information and semantic information that are stored in each pixel. Results from the suggested model were respectable.

Paper [6] proposes a CNN with a feed-forward pass and evaluate the efficiency of their approach using a "colourization Turing test". Their approach successfully tricks people on 32% of the trials, which is a significant improvement over past methods. They also show how colourization, a cross-channel encoder, is a strong pretext job for self-supervised feature learning.

Paper [7] utilises VGG-16 to categorise several peanut varieties. The fully linked layers were eliminated, global average pooling layers were added, and batch normalisation layers were added in each block before the max pooling layers.

In paper [8], the dataset of MIT places is described, and uses predefined CNN models to study the performance of this dataset.

Paper [9] The researcher used transfer learning to demonstrate image classification and image prediction using the ImageNet data set in Google Collaboratory. Transfer learning models ResNet50, MobileNet, MobileNetV2, VGG16, and VGG19 were employed in this study. Google Colab notebook was utilised for the categoristion and prediction of images.

Paper [10] presents a novel DenseUnet GAN architecture is presented for colouring NIR facial images. In order to keep facial characteristics and details in the colourization results, the proposed GAN contains both Unet and DenseNet qualities. A loss function that is optimised and includes colour loss, pixel loss, and feature loss is constructed in order to address the challenges in NIR picture colourization.

Study [11] introduced a revolutionary colourization technique that uses a reference colour image and deep learning architecture. technique produces believable results even when the target image does not have obvious correspondences in the reference, making it a general solution for exemplar-based colourization.

The research suggests employing a deep neural network with a fusion layer to smoothly combine local information based on discrete picture patches with global priors calculated from the entire image. [12]. It is made by fusing global and local features, and any picture size can use it.

In order to directly use deep learning techniques, paper [13] reformulates the colorization issue and conducts research on it. It is suggested to use a post-processing phase based on combined bilateral filtering.

In study [14], they used a Scaled-YOLOv4 raising the Peak signal-to-noise ratio (PSNR) by 2.6% and enhancing the accuracy of the colourization results. The png extension performed 5.8% better than JPEG on the Learned Perceptual Image Patch Similarity (LPIPS) metric when the outcomes of coloured images with various extensions were compared.

In the paper [15], experimental work shows that deep learning and transfer learn-ing can be used to classify COVID-19 using chest X-ray imaging. With the aid of weighted sampling, reasonable high performance can be achieved while trainable parameters in the model is decreased. In the future, performance can be enhanced by using a customised CNN architecture with a comparable topology.

In paper [16], they propose a model that blends deep convolutional neural net-work that was recently constructed with features from the Inception-ResNet-v2 model that are high-level.

In [17], using a variety of computer vision techniques and approaches, Image colorisation was carried out with an emphasis on the development and effectiveness of Generative Adversarial Networks (GANs).

[18] suggested CNN architecture made use of a number of convolutional layers and also max pooling layers. In the end, regression was per-formed using a Support Vector Machine (SVM).

A model that has previously been trained and made accessible by Google is Inception ResNet V2. At the fusion layer, features retrieved by the feature extractor fuse with the encoder's output in [19].

Paper [20] outlines the ideas that underlie existing colourization techniques, classifies them, and shows their benefits and drawbacks.

Table 1. Literature Comparison Table

Author	Method	Results
[1]	KNN, MLP, SVM	Method improves performance on measure from 0.235 to 0.823.
[2]	VGG-16	-
[3]	CNN and Quaternion Structural Similarity (QSSIM)	More accurate than Dahl
[4]	MobileNetV2, VGG16, VGG19 and ResNet50	Mobile net V2 produces more vibrant images and resnet50 is more accurate
[5]	CNN	GAN produces more accurate results (images).
[16]	Inception-ResNet-v2	Worked well on historical images and showed acceptance by the public.
[18]	CNN-SVM	Demonstrated that the suggested system could predict the coloured image from the learning knowledge acquired during the training process.
[20]	Grayscale Image Colourization Methods	As a result of their successful integration of human input with neural network automation, user-guided neural networks are the category of colourization that the data indicate holds the most promise.

3 Architecture

We will make use of a VGG-16 model that has already been trained by applying transfer learning. Our objective is to distinguish a full-coloured image, which has three values per pixel (hue, saturation, and lightness), from a grayscale image, which only has one value per pixel (lightness). There are 256x256x3 photos in the dataset. Our outputs are 224 x 224 x 2 (the other two channels), and as a result, our inputs are 224 x 224 x 1 (the brightness channel).). The architecture of our model is described in Figure 1.

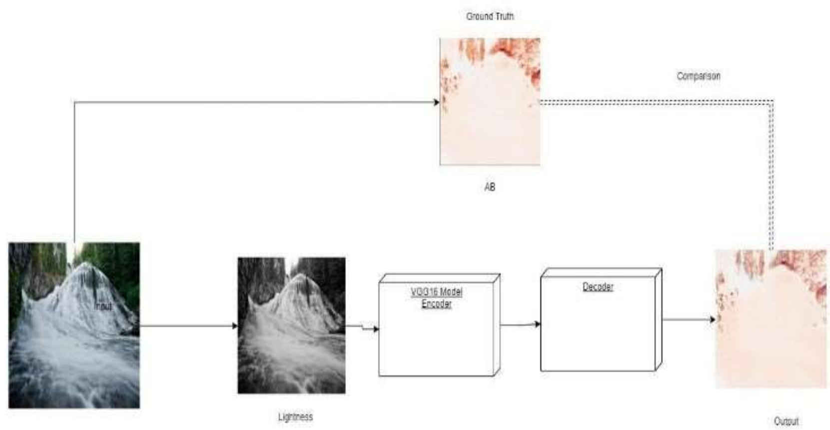


Fig. 1. Architecture of the Improved VGG-16 Model

We will work with photographs in the LAB colourspace (Lightness, A, and B), rather than the common RGB format, as people typically do. The information in this colourspace is identical to that in RGB, but it will be simpler for us to distinguish between the brightness channel and the other two (which we call A and B) [2, 5, 6].

To further understand the colourization issue, let's define it in terms of the CIE Lab colour space which is a 3-channel colour space, like the RGB, except only the a and b channels are utilised to encode colour information. L (lightness) channel data are exclusively encoded in terms of intensity [2, 6].

Our goal is to identify the a and b components of the grayscale image that we want to colour. You might think of the grayscale picture as the image's L-channel in the Lab colour space. By utilising standard colour space transformations, the lab picture that was so acquired may be converted to the RGB colour space [6, 7].

Calculations are made easier by quantizing the ab space into 313 bins. The a and b values for each pixel may be replaced with a bin number between 0 and 312 thanks to quantization. The L channel, which allows values between 0 and 255, is one way to approach the problem. However, the ab channel, which accepts values between 0 and 312, has to be found. With 313 potential classes for each grey pixel, the colour prediction challenge has been changed into a multinomial classification problem [6].

3.1 Encoder

We first apply a number of convolutional layers to extract features from our image before applying deconvolutional layers to upscale (increase the spatial resolution of our features). Our encoder is made up of these. In a traditional VGG-16 model, the input layer is constructed in such a way that the first layer is formed by 3 X 3 kernel and consists of 64 feature kernel filters [5]. Our version of the Improved VGG-16 has been described in Figure 2.

Here is the fine tuning we will do in order to use VGG-16 for our dataset:

1. The initial layer of the network will be changed so that it only receives grayscale input as opposed to colourful data.
2. We will only use VGG-16 till the fifth pooling layer.
3. Since we are doing regression, we will use a mean squared error loss function. This function seeks to minimise the squared distance between the colour value we are attempting to predict and the actual (true) colour value. With the Adam optimizer, we will improve our loss function (criterion)..

The improvements we will add to a traditional VGG-16 model in order to improve its accuracy are:

1. We will be adding a Batch Normalization layer after every Max Pooling Layer. A comprehensive approach of parametrizing almost any deep neural network is provided by batch normalisation. Reparameterization greatly reduces the difficulty of updating plans at various levels. Batch normalisation improves network training. Batch normalisation allows us to employ much higher learning rates, which speeds up network training even more. It Internal covariant shift shrinks. reduces gradients' dependency on the size of the parameters or their underlying values [7].
2. We will be adding a Dropout Layer at the end of the model. Nodes in a neural network are dropped out when a phrase like "dropout" is used. The parent network's forward and backward links with dropped nodes are momentarily removed to create a new network architecture. A unit may alter during overfitting in a way that corrects the flaws of the other units. This causes complicated co-adaptations, which in turn causes the overfitting issue because if we employ dropout, it prevents these units from correcting the error of other units, inhibiting co-adaptation. This complex co-adaptation fails to generalise on the unseen dataset.It precludes the synchronised weight optimisation of all neurons in a layer. The weights are decorated as a result of this adaptation's random grouping, which prevents all neurons from convergent towards the same objective [6].
3. We will also perform Image Augmentation on the dataset. The kinds of Image Augmentation we are going to use are horizontal flip, zoom in and zoom out, as well as changing brightness. Using existing data, image augmentation makes changed dataset copies that are then used to expand the training set. A few minor changes must be made to the dataset. It improves the M L model's performance while reducing the operating expenditures related to data collection. By including fresh and unique instances in training datasets, data augmentation enhances the functionality and output of machine learning models. A large and sufficient dataset is necessary for a machine learning model to function more effectively and precisely. Additionally, it lowers the price of data collection and labelling.

Figure 2 shows the layers of our Improved VGG-16 model.

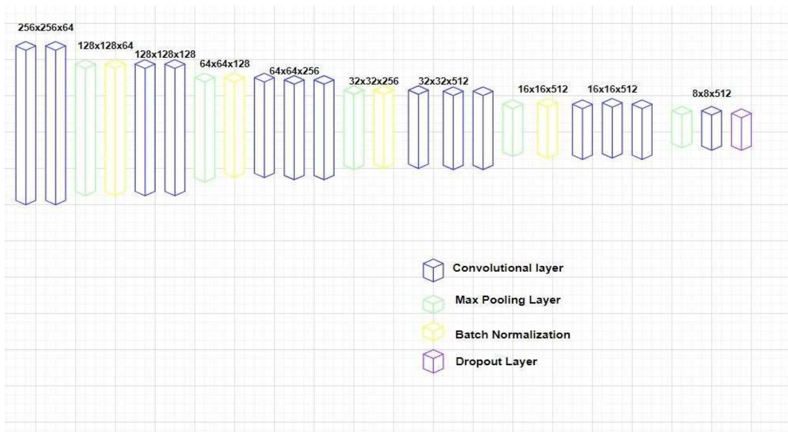


Fig. 2. Layers used in the VGG-16 model.

3.2 Decoder

The decoder uses Convolutional layers and UpSampling Layers to output the image in the correct dimensions of 256x256. While training, the coloured images is split into 2- one containing the Lightness channel, which would be grayscale, and one containing information about the colours, and the ab channels. The 'colour' image is used as Ground Truth to compare the output with, and therefore, train the model.

4 Result

The dataset used by us is the MIT Places dataset which has over 400 categories and 10 million images. It has been used in a variety of studies for image colourization [8]. Some noticeable results could be seen after the implementation of our improvements. The colours were placed more accurately. Batch Normalization and Dropout both deal with the problem of overfitting that VGG-16 is prone to, and Image Aug-mentation helps increase robustness. Together, they increase the precision of the default VGG-16 model. Figure 3 shows the difference between a traditional VGG-16 model, the use of Batch Normalization and Dropout layers, the use of Image Aug-mentation, as well as the Improved Model, which uses all three.

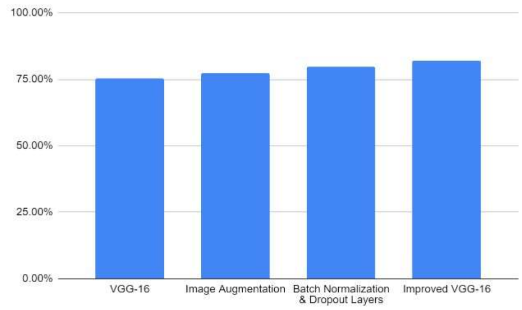
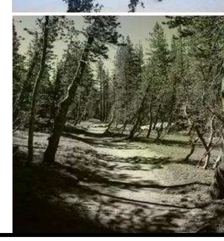


Fig. 3. Comparing accuracy at different optimizations

The improved model is able to predict colour better than the normal VGG-16 model. Table 2 shows the output from a standard VGG-16 model, and one from our improved model.

Table 2. Comparison of the output from the standard and the improved VGG-16 model

VGG-16	Improved VGG-16
	
	

5 Conclusion

Image Colourization is a hard task, and at this stage, requires human intervention to be 100% successful. However, with the use of Aritificial Intelligence, both time and effort can be reduced. On comparison, our model works better than most models based on VGG-16.

In the future, more work can be done on how to best use both Batch Normalization layers and Dropout layers in order to increase accuracy and avoid overfitting. Other architectures, such as the ones based on hyper columns and using classifier models can also be used with the suggested optimizations.

References

1. An, J., Gagnon, K. K., Shi, Q., Xie, H., & Cao, R. (2019, October). Image colourization with convolutional neural networks. In 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-4). IEEE.
2. Dabas, C., Jain, S., Bansal, A., & Sharma, V. (2020). Implementation of image colourization with convolutional neural network. International Journal of System Assurance Engineering and Management, 11(3), 625-634.
3. Varga, D., & Szirányi, T. (2016, December). Fully automatic image colourization based on Convolutional Neural Network. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 3691-3696). IEEE.
4. Shaha, M., & Pawar, M. (2018, March). Transfer learning for image classification. In 2018 second international conference on electronics, communication and aerospace technology (ICECA) (pp. 656-660). IEEE.
5. Kiani, L., Saeed, M., & Nezamabadi-Pour, H. (2020, September). Image colourization using a deep transfer learning. In 2020 8th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS) (pp. 027-032). IEEE.
6. Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colourful image colourization." Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14. Springer International Publishing, 2016.
7. Yang, H., Ni, J., Gao, J., Han, Z., & Luan, T. (2021). A novel method for peanut variety identification and classification by Improved VGG16. Scientific Reports, 11(1), 1-17.
8. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6), 1452-1464.
9. Gujjar, J. P., Kumar, H. P., & Chiplunkar, N. N. (2021). Image classification and prediction using transfer learning in colab notebook. Global Transitions Proceedings, 2(2), 382-385.
10. Xu, J., Lu, K., Shi, X., Qin, S., Wang, H., & Ma, J. (2021). A DenseUnet generative adversarial network for near-infrared face image colourization. Signal Processing, 183, 108007.
11. He, M., Chen, D., Liao, J., Sander, P. V., & Yuan, L. (2018). Deep exemplar-based colourization. ACM Transactions on Graphics (TOG), 37(4), 1-16.
12. Hussain, AS Sahil, et al. "Automatic Colourization and Restoring of gray scale images using Deep Learning." Think India Journal 22.10 (2019): 2473-2479.
13. Cheng, Zezhou, Qingxiong Yang, and Bin Sheng. "Deep colourization." Proceedings of the IEEE international conference on computer vision. 2015.
14. Hesham, Mennatullah, Heba Khaled, and Hossam Faheem. "Image colourization using Scaled-YOLOv4 detector." International Journal of Intelligent Computing and Information Sciences 21.3 (2021): 107-118.
15. Kumar, S., & Kumar, H. (2023). X-Ray Image Classification of Covid-19 Using Transfer Learning with Vgg16 and Vgg19. Available at SSRN 4359652.

16. Federico Baldassarre, Diego González Morín, Lucas Rodés-Guirao Deep Koalarization: Image Colourization using CNNs and Inception-ResNet-v2
17. Rashi Dhir, Meghna Ashok,Shilpa Gite AN OVERVIEW OF ADVANCES IN IMAGE COLOURIZATION USING COMPUTER VISION AND DEEP LEARNING TECHNIQUES
18. Abualola, A., Gunawan, T.S., Kartiwi, M., Ambikairajah, E., Habaebi, M.H. (2021). Development of Colourization of Grayscale Images Using CNN-SVM. In: , et al. Advances in Robotics, Automation and Data Analytics. iCITES 2020. Advances in Intelligent Systems and Computing, vol 1350. Springer, Cham. https://doi.org/10.1007/978-3-030-70917-4_6
19. R. S. Shankar, G. Mahesh, K. V. S. S. Murthy and D. Ravibabu, "A Novel approach for Gray Scale Image Colourization using Convolutional Neural Networks," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-8, doi: 10.1109/ICSCAN49426.2020.9262377.
20. Žeger, S. Grgic, J. Vuković and G. Šišul, "Grayscale Image Colourization Methods: Overview and Evaluation," in IEEE Access, vol. 9, pp. 113326-113346, 2021, doi: 10.1109/ACCESS.2021.3104515.