Int J Syst Assur Eng Manag (June 2020) 11(3):625–634
https://doi.org/10.1007/s13198-020-00960-5

ORIGINAL ARTICLE

Implementation of image colorization with convolutional neural network

Chetna Dabas[1] · Shikhar Jain[1] · Ashish Bansal[1] · Vaibhav Sharma[1]

Received: 16 November 2018 / Revised: 9 December 2019 / Published online: 6 March 2020

Abstract Huge amount of work is getting done on Image colorization worldwide. This research paper proposes a model for image colorization while making use of fully automatic Convolutional Neural Network. Image colorization processes a daunting task, and this research paper proposes a relevant model for the prediction of A and B models for LAB color space and it makes a direct use the lightness channel. In this work, a pre-trained VGG-16 network was used for semantically interpreting the concepts associated with images and coloring the images. In the proposed work, the convolutional layer has been fused with the max pooling layer (higher one) of the VGG network. Architecture of the proposed model has been presented. The experimentation has been carried out with varying objective functions. LaMem experimental dataset has been used in this work in order to validate the proposed model. The proposed model is evaluated and results are visualized by histograms for true and predicted images for RGB values. Further, the proposed model has been compared with the existing models and performs better in terms of execution times (in s) for different image sizes and the results are tabulated.

Keywords Image colorization · Convolutional neural network · VGG-16 model

1 Introduction

Automatic colorization deals with coloring grey scale images without any direct human assistance. Image colorization has many applications ranging from coloring historical images to improvement in surveillance feed. This problem also facilitates solution to many other problems that require changes in pixel values in training images to attain some desired results.

With so many researches going on image colorization topic, this problem still considers to be a complex one as colorization of images takes more than just knowing what is in image. As an example a car can be colored red or green. A Tree can be colored green and same tree can be colored brown depending on image is captured in spring season or autumn season.

Recent researches in field of computer science and image processing show that CNN has emerged as a leader in image processing, achieving responsive solutions to complex image problems with least error rate compared to other methodology. CNN has much better success rate in learning and differentiating between colors patterns and shapes in images as compared to other traditional methods. In image classification using CNN, the research community has achieved makeable results (Krizhevsky et al. 2012), handwritten character classification (Cireşan and Meier 2015) using CNN.

A reliable solution would comprise identifying most matching pixels from the images and referring it to the target pixel for colorization, hence producing a much better

✉ Chetna Dabas
chetna.dabas@jiit.ac.in

Shikhar Jain
15103105shikhar@gmail.com

Ashish Bansal
ashish15103121@gmail.com

Vaibhav Sharma
vaibhav15103136@gmail.com

[1] Department of CSE&IT, Jaypee Institute of Iinformation Technology, Noida, India

result and faster. The base of our model is comprised up of two neural architectures namely CNN and VGG-16 (Simonyan and Zisserman 2014).

The VGG-16 model was the runner-up at the image-net challenge [ILSVRC-2014] and provided a base for future image net competition as it is highly accurate and has a linear structure as compared to other image-net modals for better visualization and structure.

The organization of this research paper is as follows: Related work is described in Sect. 2, experimental datasets utilized in the proposed research work is explained in Sect. 3, further, the empirical aspects related to the proposed approach is depicted in Sects. 4 and 5 describes the experimentation and results carried out as a part of this research work and finally Sect. 6 concludes with conclusions with future work.

2 Related work

In the existing literature, in recent past few years, a lot of work has been carried out by the researchers in the image colourization domain with lots of applications. To name a specific application in the similar domain, a very recent work one of such application has been investigated by the authors (Ahmad et al. 2018) while examining image cryptosystem in order to secure coloured images during the process of transmission. These authors have analyzed the security aspects in image colorization domain. In Bala and Kaur (2016), the authors proposed an algorithm to produces a codebook which is efficient and it has less computational time, the results also gives good PSNR Work so far on image colorization can be classified into three classes. These classifications differ in the way they work on the coloured images and grey scale images. The classifications are Learning-based colourization, Scribble-based colorization and Example-based colorization respectively.

Scribble-based colorization was introduced by Levin et al. (2004). This work is also known as colour optimization. In this technique, the user places the scribbles on the gray-scale target image. After that, the colour information is bonded with the corresponding image pixels. This happens all over the image. Huang et al. (2005) later improved it by optimizing the image edges and reducing image blending. Later Yatziv and Sapiro (2006) in his paper Video colorization and fast image by making use of chrominance blending improved colorization by combining the scribble colors and defining the most appropriated color for the pixel, in order to minimize the distance between the scribble and the pixel, a distance matrix was computed and weights were defined on the basis of distance metric. The method required more of user interaction and was not automatic colorization.

Example-based colorization another classical approach for colorization where the user is provided with an example image unlike scribble, the user does not need to have experience with colorization. It was introduced by Welsh et al. (2002) who developed his modal on colorization by the efforts of Hertzmann et al. Image Analogies and Reinhard et al. color transfer technique, made a model which colorizes a pixel based on the swatches between the target image and the example image. It was successful in targeting the swatches and predicted the color tune accurately. Later Charpiat et al. (2008) modified the model by adding a top layer where similar colours are used to colour the regions with similar textures. The modal performed well but these methods were limited heavily on the quality of the reference example image provided by the user and can be a difficult task to find such images. The most recent example-based model was by Aditya Deshpande, Jason Rock and David Forsyth, they built an objective function with coefficients related to image features and then minimizes its coefficients.

Learning-based colorization Comprises of patch based colorization which was brought by Bugeau and Ta (2012) which takes square patches around each pixel. The patches data on luminance were extracted and distance selection strategy was proposed to train the model. Later a three-layer deep network was proposed by Cheng for fully automatic colorization, the three layers of data was extracted from each pixel and concatenated to train the color model. The three layers are grey scale values, daisy values and semantic features. The layers values were extracted and feed into the neural network for training, his model predicted RGB values which produced inaccurate results.

In a related research paper (Cheng et al. 2015), the authors investigate the problem of image colorization in order to transform a greyscale image into a colourful one. Differently, unlike the previous techniques, this paper used fully automatic method for colorization while exploiting deep coloring techniques and modelling of large scale data.

Unlike the previous methods, this paper aims at a high-quality fully-automatic colorization method. With the assumption of a perfect patch matching technique, the use of an extremely large-scale reference database (that contains sufficient color images) is the most reliable solution to the colorization problem. However, patch matching noise will increase with respect to the size of the reference database in practice. Inspired by the recent success in deep learning techniques which provide amazing modeling of large-scale data, this paper re-formulates the colorization problem so that deep learning techniques can be directly employed. To ensure artifact-free quality, a joint bilateral filtering based post-processing step is proposed. Numerous experiments demonstrate that our method outperforms the state-of-art algorithms both in terms of quality and speed.

2.1 More applications

A recent research work (Yan et al. 2018) reports that the encoding of alike images must be mapped to alike binary codes and the other way round. They also mentioned that there must be minimization of quantization loss with respect to Hamming and Euclidean spaces and the there must be an even distribution of the learned codes. The authors proposed a method for the improvement of power (discriminative) of the associated binary codes. They have also carried out comparisons of their proposed method with the existing algorithms.

The authors of a recent research work (Chen et al. 2018), discusses the convolution for dense predicted tasks carried out taking un sampled filters into consideration. The authors highlights that there exists a toll in terms of the localization accuracy when a collection of down sampling and max-pooling is considered in the deep Convolutional neural networks although invariance is achieved during the process. The authors of this work have addressed this problem by the combination of responses which are retrieved from final deep Convolutional neural network which posses a conditional random field that is fully connected. Their proposed “DeepLab” framework initiates a PASCAL VOC-2012 semantic image segmentation process, with 79.7 percent mIOU associated with the test set. The authors of this work claims that there is a significant enhancement in the results for three other existing datasets namely PASCAL-Person-Part, Cityscapes and PASCAL-Context.

In a recent research work (Esteva et al. 2017), deep neural networks have been used for the classification of skin cancer using artificial intelligence techniques. The authors of this research work claims that the results retrieved as a part of their work are competent enough to be adopted by the dermatologists in the identification of both common as well as deadliest skin cancer.

In the proposed research work, the implementation of the learning-based colorization is carried out; the proposed model is heavily inspired by Jeff Huang and You Zhou deep CNN automatic colorization system (Hwang and Zhou 2016). Jeff’s model relies on VGG-16 pre-trained Image Net layers whereas in the proposed modal the authors only fused the higher layers of the VGG-16 modal.

In one recent research paper (Quan et al. 2019) a particular forensic problem has been picked up to distinguish between natural and colorized images and investigates and observations are addressed along with the results. Moreover, these autors have proposed a technique to colorized image detection performance while considering various settings at the first network layer and merging the decision results from Convolutional neural network architectures.

In yet another recent work (Jiang et al. 2019), the authors have presented a method for the process of fully automatic image colorization process. This makes use of CNN for the analysis of relationships which are non- linear amongst the chrominance points and image features. For the extraction of high level image features, a pre-trained residual network is also utilized in this work.

3 Experimental datsets used

The proposed model was tested on several datasets. The MIT LaMem dataset used while experimentation. This dataset consists up of 58,741 images developed from various existing images. In the proposed work, experimentation with larger datasets is carried out. In specific, the MIT LaMem Database contains a large number of images as seen in Fig. 1.

The dataset under consideration holds object-centric images, scene-centric images, and other images which are capable of igniting certain emotions. Apart from this, it contains user captured images like ‘selfies’. In this work, the exploration of the correlation amongst a large variety of attributes has been made. Due to this consideration of the large variations of images made LaMem dataset specifically suitable for the training of the proposed colour model.

4 Proposed approach

The proposed model is developed that comprises image pre-processing and neural network feeding of images. Various aspects of the proposed model are explained and presented in the sub-sections ahead:

4.1 Preprocessing

In this work, a dataset of 58,000 images and each image was of different dimensions was considered. So, first each image was resized in 128 * 128 * 3 pixels. Now, our images were in RGB format. But, there is a problem with RGB format. If we train our model in this format, then we have to predict 3 color channels for each image. This would increase complexity of our model. Instead of doing that, we transformed each image into LAB color space.

The Lab color space dictates all perceivable colors (mathematically) in the three dimensions A and B for colour components, L is for lightness. L can be represented as follows:

Fig. 1 Few samples from LaMem dataset (http://memorability.csail.mit.edu/explore.html), showing the diversity of images like selfies, arts and architects

$$L = \frac{(R + G + B)}{3} \tag{1}$$

The Lab color space is represents human vision (approximate). This color space was utilized so as to reduce the number of variables in order to predict down to two.

4.2 Architecture

Our model architecture is inspired by Ryan's model (http://memorability.csail.mit.edu/explore.html) with some modifications. We used a pre-trained VGG-16(Simonyan and Zisserman 2014) network with modifications. This VGG model identifies objects in images. This will give our model semantic information about the images. As our training images are of different size so we resized them to 128 * 128 * 3, then to remove any non-linearity from training images we performed basic conv_2d operation with stride of [1, 1, 1, 1] and same padding followed by 'Relu.'

The output after applying 'Relu' served as input to pre-trained convolutional neural network based model which perform multiple convolution (down sampling) followed by multiple 'Relu' and 'max pool' operations through different layer of its architecture in order to extract features from the training/testing images. In the process we stored output of each layer of VGG-net which helped us later to generate segments/colored image from greyscale image.

Now, our task is to regenerate images from features which was extracted formerly, so in order to do that we first apply transposed convolution (up sampling) on final output (5th pooling layer) from model mentioned above. The 'new output' (after up sampling) generated has same dimensions as that of output of fourth layer of pre-trained model.

Now we fused original output of fourth pooling layer of VGG model and the output obtained after up sampling of output of final layer in order to predict our own result for fourth layer, then we again apply fusion on above fused output with the original output from third layer of VGG network to predict result for second layer. Finally, we applied transposed convolution on our predicted output (of second layer) to resize our predicted image to the size of original image i.e. 128 * 128 * 3. As shown in Fig. 2.

Architectural change made by us is that we only fuse output of higher level layer because these layer capture more concepts and main features of the image as compared to lower level layer which captures only tone and texture of the image. Fusing only high layers reduces complexity of our model.

4.3 Loss functions

We tried different loss functions to converge our model early. First We tried L1 loss function. L1 loss function is also known as least absolute error (LAE). It basically minimizes the sum of pixel differences between Predicted image and True image. S can be represented as follows:

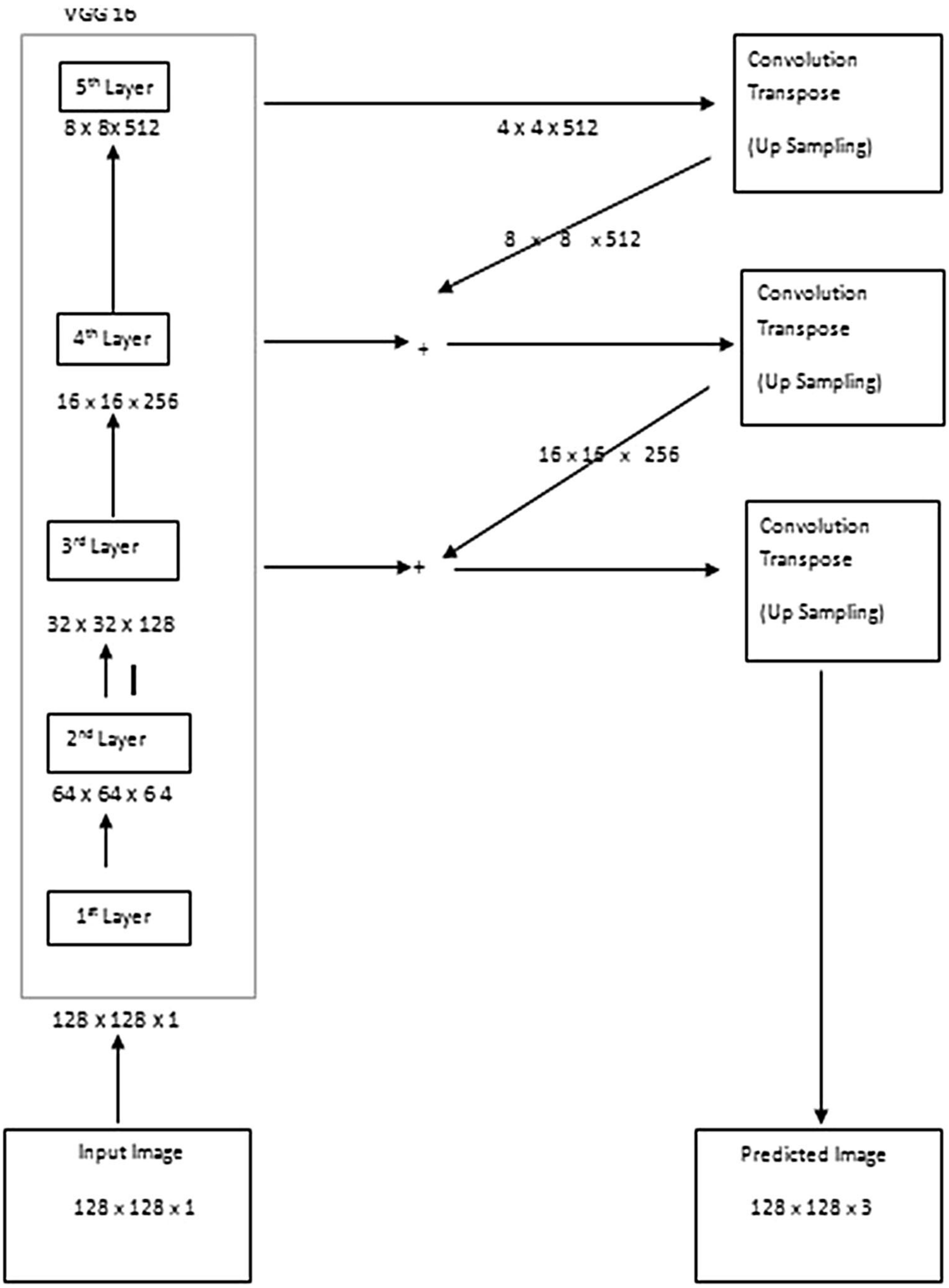


Fig. 2 Architecture of the proposed model

$$S = \sum_{i=1}^{n} |y_i - f(x_i)| \tag{2}$$

Here Yi is basically pixel matrix of true image and f(Xi) is pixel matrix for predicted image. We tried to find minima of this error function using Adam Optimizer with learning rate 1e-4. We get pretty decent results with this loss function.

Then, we tried L2 loss function in hope to converge our loss function earlier. L2 loss or least square error function calculates the square difference between pixel matrices of true and predicted image.

$$S = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{3}$$

As expected, this loss function converges fast as compare to l1 loss function using Adam optimizer with same learning rate.

But, during training the authors found out that L1 and L2 loss function tries to average out the difference of pixels in predicted and true image and the reduction in sharpness of the images. This also results in color patches in between image. For images like cars which can take any color it tries to average out the color and result is it tries to color greyish the image.

To counter this problem, we used cross entropy like function. We defined our loss function as

$$L(AB_p, AB) = \left(-\frac{1}{W * H}\right) \sum_{n=1}^{n} (AB) * \ln(AB_p) \quad (4)$$

Here $\mathbf{AB_P}$ is predicted value of A and B color space values and **AB** is A and B color spaces of true image. H is the height and W is the width of the image under consideration. This loss function gives better results than L2 loss function. We verified the results using SSIM (structural similarity matrix) as explained in later section.

4.4 Proposed architecture

Figure 2 presents architecture of the proposed model which is self explanatory. The input to this model is a selected image q with only L component. The output is the q image with 3 components namely L, A and B.

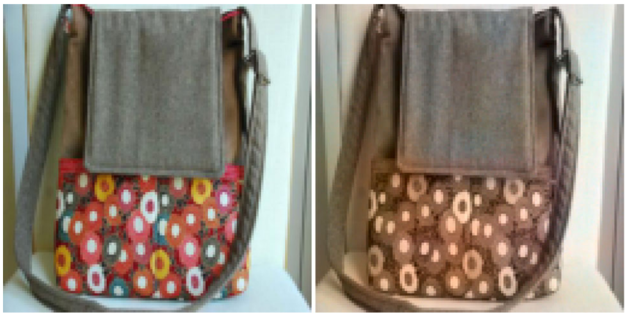


Fig. 4 Showing the limitations of colorization as the model averages out the color

4.5 Motivation and contribution

The proposed method is applicable to a large scale reference image dataset. The use of Convolutional neural network in the specified architecture supports the combination of diverse features associated with pixel. It also computes the associated chrominance points. Moreover, the exiting state

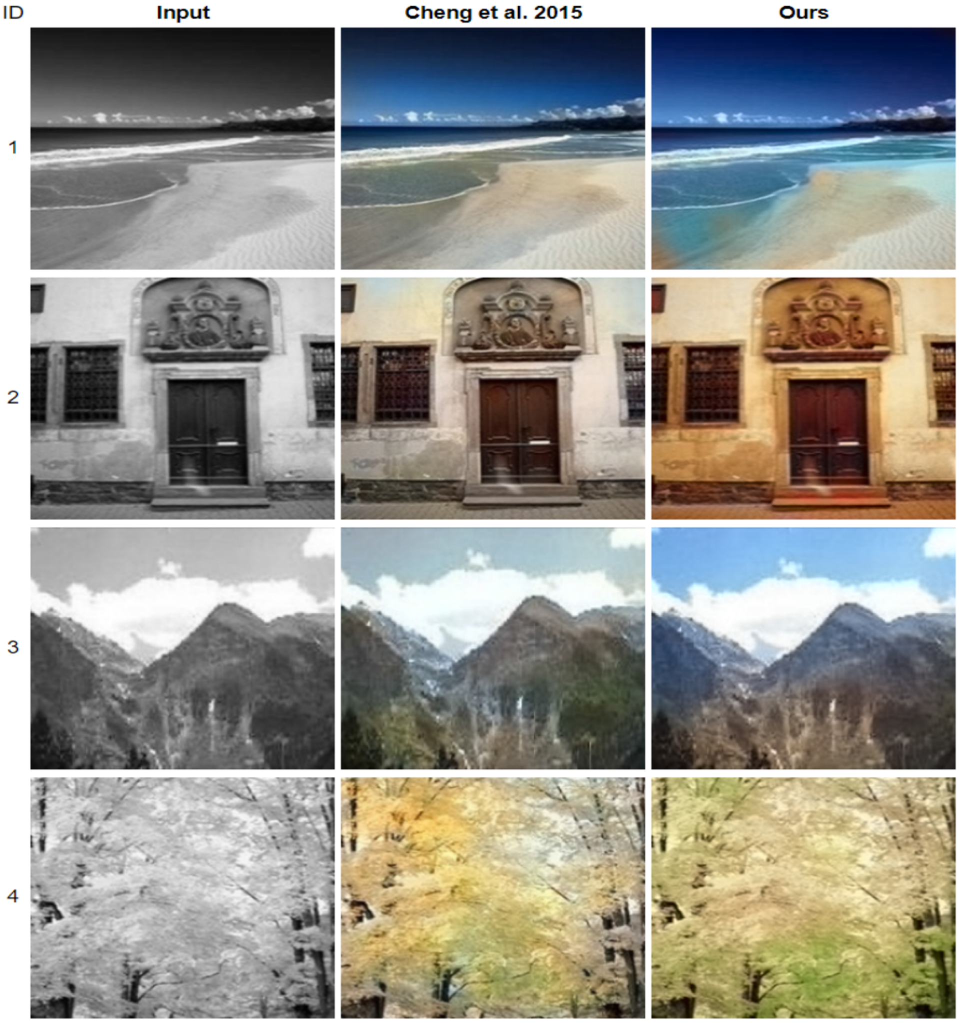


Fig. 3 Comparison of proposed model with Cheng et al. (2015) with grey scale image as input

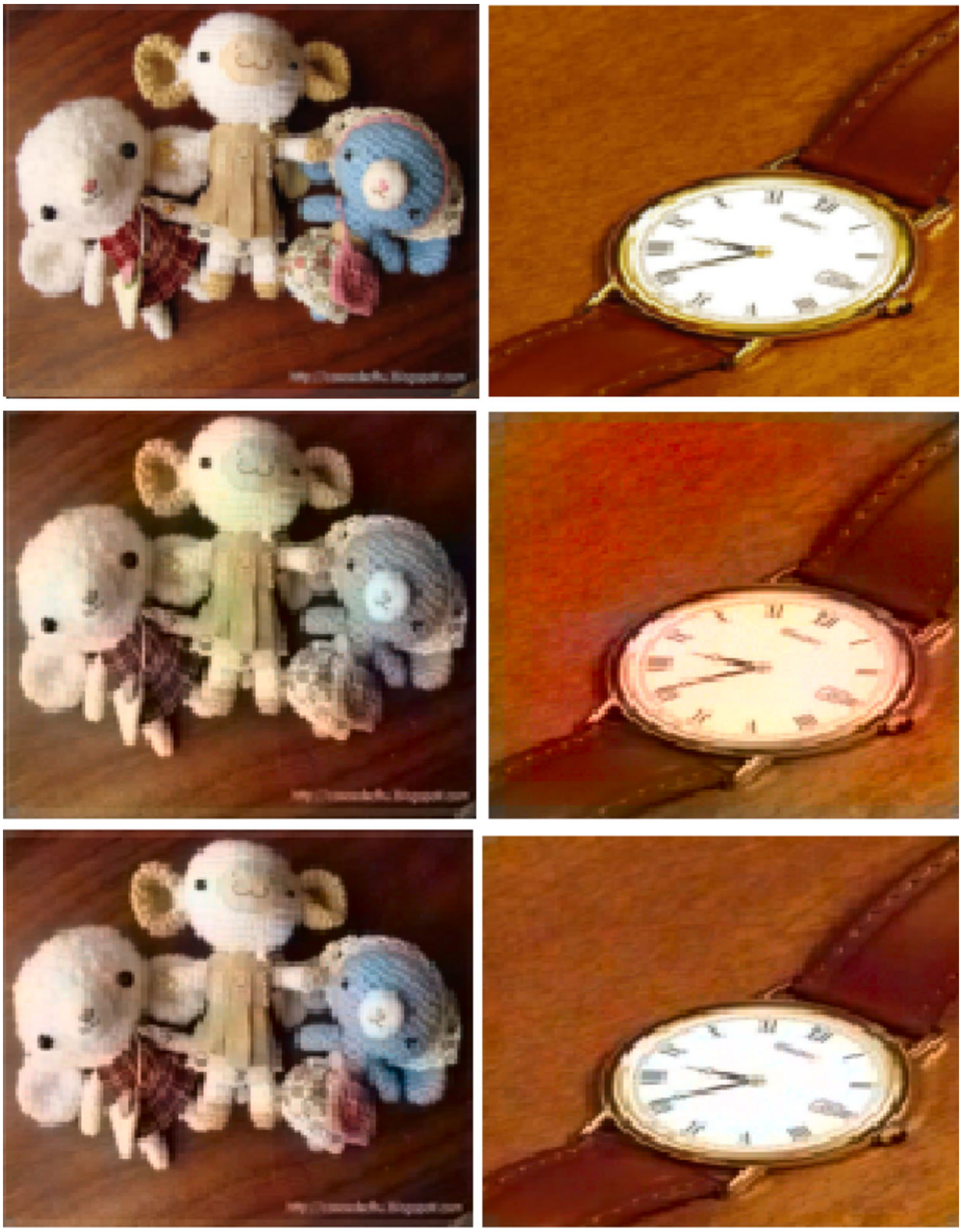


Fig. 5 Original (top), L2 predicted (middle) and cross entropy predicted (bottom)

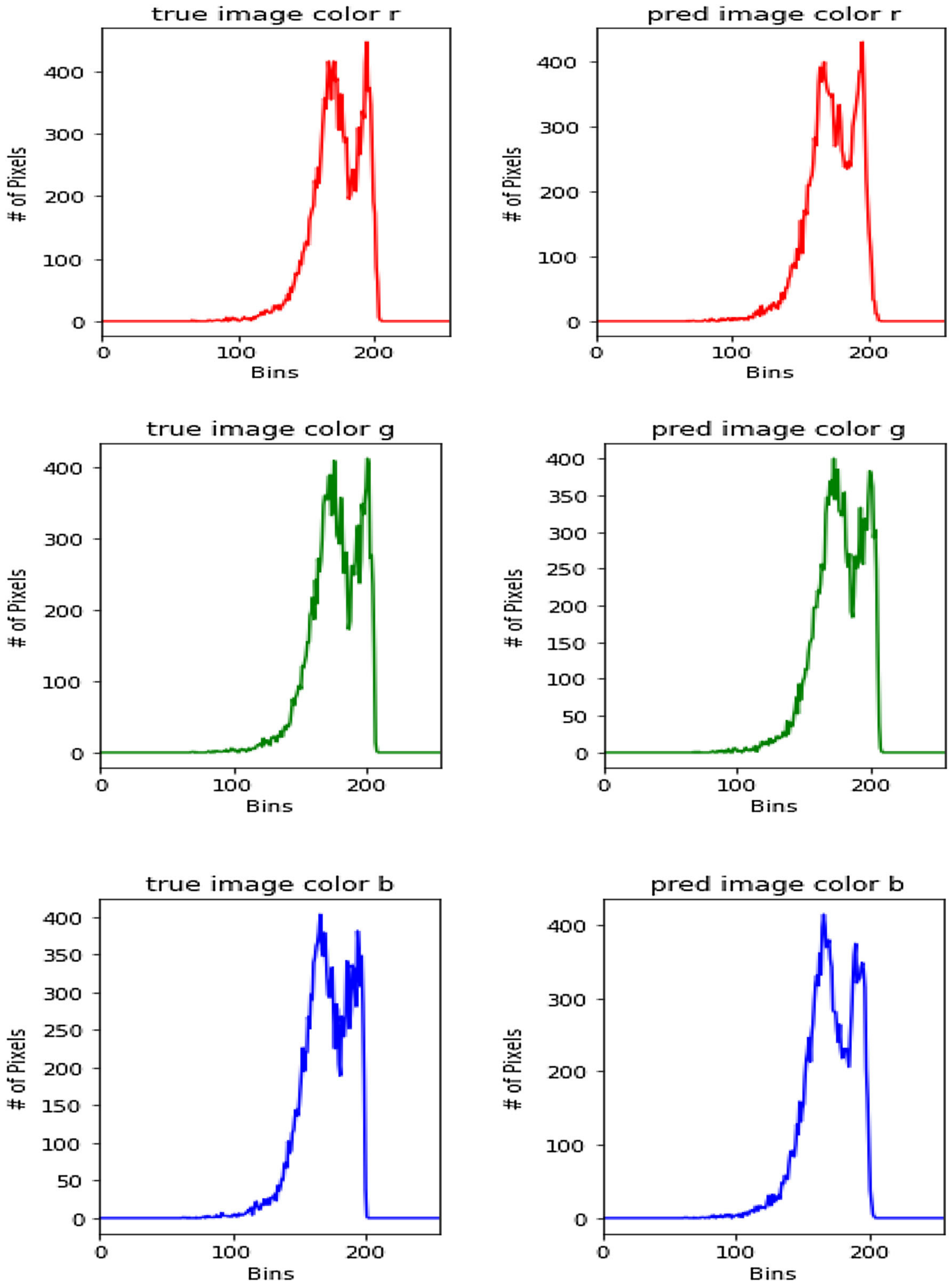
of art techniques is comparatively slow due to the reason that they keep on computing the super pixel values from huge participants whereas the proposed method with targeted architecture is trained to handle large scale data efficiently.

5 Experimentation and results

In the proposed work, the full 58,741 images were used for training from the La Mem dataset (http://memorability.csail.mit.edu/explore.html). The authors compared the proposed model in Fig. 3. With the Cheng et al. (Cheng et al. 2015) work, its quantitative comparison was not possible as he did not release his code so we processed the images which the authors retrieved from cropping (manually) from the Cheng et al. paper.

Cheng et al. work (2015) investigates the problem of colorization and is responsible for converting a grayscale image into a colorful image and here a post processing step joint bilateral filtering is initiated.

The experimental results presented in this research work corresponding to the proposed model standout from him and provided a better understanding of images. Cheng model was an example-based colorization and utilised several images to simultaneously to predict the result whereas our model is a learning-based colorization which utilises CNN to predict the result.


true image color r

pred image color r

true image color g

pred image color g

true image color b

pred image color b

Fig. 6 Histograms of true colour versus predicted colour (For red, green and blue colours) (color figure online)

The result snapshot in Fig. 4 depicts the images of colourization as the model averages out the colour.

As shown in Fig. 4, the proposed model averages the color in different colored circles. So, to color multi-colored objects use of GAN (Generative Adversarial Networks)

Table 1 Comparison of the execution times (s) of images of different resolutions (proposed versus existing models)

Algorithm	(256 × 256)	Execution times (s) (512 × 512)	(1024 × 1024)
Proposed	2.146	4.213	31.145
Cheng et al. (2015)	6.780	17.423	63.142
Qu et al. (2008)	251.709	712.149	5789.075

was made which uses encoder to create false image and decoder to minimize the difference between false image and true image.

We evaluated our model on three loss functions L1, L2 and Cross entropy. The results of L2 and cross entropy differed in terms of saturation and sharpness of colours as seen in Fig. 5, predicted image of toy using L2 loss function has faded blue colour whereas cross entropy loss function image has bright colours.

Similarly the watch image coloured using l2 loss function has random red patch whereas cross entropy loss function predicted saturated colours in image.

Figure 6 presents improved result histograms comparisons for the coloured component distribution for the proposed model. The result histograms in this figure are corresponding to red, green and blue values of true versus predicted image. The visualization of colour distribution is achieved in Fig. 6. The left graph in the each of these snapshots depicts true image colour and the right graph depicts the predicted image. Table 1 presents a comparative view of execution times in seconds of images with varying sizes of the proposed model with existing ones.

6 Conclusions and future work

This research paper proposed a fully automatic and novel approach to image colorization based on VGG-16 and Convolution Neural Network. This research work also threw light on the how CNN can be applied in domain of image colorization and the way it can be used in future models. The VGG-16 pre-trained model was used to fetch the semantic information and image characteristics of an image which was processed in the CNN architecture. The A and B color were predicted of the LAB image layers and where L is Intensity layer which remains unchanged. The proposed model has been compared with existing models (Qu et al. 2008; Cheng et al. 2015) and, it is concluded by experimental results that the fully automatic proposed model performs better for execution times (in seconds) than the existing methods for varying large scale reference image sizes. The proposed model holds potential of getting accepted in media industries across the globe.

References

Ahmad M, Doja MN, Beg MS (2018) Security analysis and enhancements of an image cryptosystem based on hyperchaotic system. J King Saud Univ-Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2018.02.002

Bala A, Kaur T (2016) Local texton XOR patterns: a new feature descriptor for content-based image retrieval. Eng Sci Technol Int J 19(1):101–112

Bugeau A, Ta VT (2012) Patch-based image colorization. In: 21st international conference on pattern recognition (ICPR), IEEE, pp 3058–3061

Charpiat G, Hofmann M, Schölkopf B (2008). Automatic image colorization via multimodal predictions. In: European conference on computer vision, Springer, Berlin, pp 126–139

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE Trans Pattern Anal Mach Intell 40(4):834–848

Cheng Z, Yang Q, Sheng B (2015) Deep colorization. In: Proceedings of the IEEE international conference on computer vision, pp 415–423

Cireşan D, Meier U (2015) Multi-column deep neural networks for offline handwritten Chinese character classification. In: 2015 International joint conference on neural networks (IJCNN), IEEE, pp 1–6

Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115

Huang YC, Tung YS, Chen JC, Wang SW, Wu JL (2005) An adaptive edge detection based colorization algorithm and its applications. In: Proceedings of the 13th annual ACM international conference on multimedia, ACM, pp 351–354

Hwang J, Zhou Y (2016) Image colorization with deep convolutional neural networks. In Stanford University, Technical Report

Jiang, H., Tang, S., Li, Y., Ai, D., Song, H., & Yang, J. (2019). Endoscopic image colorization using convolutional neural network. In: 2019 IEEE 7th international conference on bioinformatics and computational biology (ICBCB), IEEE, pp 162–166

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Jordan MI, LeCun Y, Solla SA (eds) Advances in neural information processing systems. MIT Press, Cambridge, pp 1097–1105

Levin A, Lischinski D, Weiss Y (2004) Colorization using optimization. In: ACM transactions on graphics (tog), ACM, vol 23, no 3, pp 689–694

Margulis D (2005) Photoshop LAB color: The canyon conundrum and other adventures in the most powerful colorspace. Peachpit Press, Berkeley

Qu Y, Pang WM, Wong TT, Heng PA (2008) Richness-preserving manga screening. In ACM transactions on graphics (TOG), ACM, vol 27, no 5, p 155

Quan W, Wang K, Yan DM, Pellerin D, Zhang X (2019). Impact of data preparation and CNN's first layer on performance of image forensics: a case study of detecting colorized images. In: IEEE/

WIC/ACM international conference on web intelligence, ACM, vol 24800, pp 127–131

R. Dahl image colorization http://tinyclouds.org/colorize

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

Welsh T, Ashikhmin M, Mueller K (2002) Transferring color to greyscale images. In: ACM transactions on graphics (TOG), ACM, vol 21, no 3, pp 277–280

Yan C, Xie H, Yang D, Yin J, Zhang Y, Dai Q (2018) Supervised hash coding with deep neural network for environment perception of intelligent vehicles. IEEE Trans Intell Transp Syst 19(1):284–295

Yatziv L, Sapiro G (2006) Fast image and video colorization using chrominance blending. IEEE Trans Image Process 15(5):1120–1129

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.