Perceptual Conditional Generative Adversarial Networks for End-to-End Image Colourization

Shirsendu Sukanta Halder, Kanjar De(✉), and Partha Pratim Roy

Indian Institute of Technology Roorkee, Roorkee 247667, India
shalder@cs.iitr.ac.in,{kanjar.cspdf2017,proy.fcs}@iitr.ac.in

Abstract. Colours are everywhere. They embody a significant part of human visual perception. In this paper, we explore the paradigm of hallucinating colours from a given gray-scale image. The problem of colourization has been dealt in previous literature but mostly in a supervised manner involving user-interference. With the emergence of Deep Learning methods numerous tasks related to computer vision and pattern recognition have been automatized and carried in an end-to-end fashion due to the availability of large data-sets and high-power computing systems. We investigate and build upon the recent success of Conditional Generative Adversarial Networks (cGANs) for Image-to-Image translations. In addition to using the training scheme in the basic cGAN, we propose an encoder-decoder generator network which utilizes the class-specific cross-entropy loss as well as the perceptual loss in addition to the original objective function of cGAN. We train our model on a large-scale dataset and present illustrative qualitative and quantitative analysis of our results. Our results vividly display the versatility and the proficiency of our methods through life-like colourization outcomes.

Keywords: Colourization · Generative Adversarial Networks · Image Reconstruction

1 Introduction

Colours enhance the information as well as the expressiveness of an image. Colour images contain more visual information than a gray-scale image and is useful for extracting information for high-level tasks. Humans have the ability to manually fill gray-scale images with colours taking into consideration the contextual cues. This clearly indicates that black-and-white images contain some latent information sufficient for the task of colourization. The modelling of this latent information to generate chrominance values of pixels of a target gray-scale image is called colourization. Image colourization is a daunting problem because a colour image consists of multi-dimensional data according to defined colour-spaces whereas a gray-scale image is just single-dimensional. The main obstacle is that different colour compositions can lead to a single gray level but the reverse is not true.

C. V. Jawahar et al. (Eds.): ACCV 2018, LNCS 11362, pp. 269–283, 2019.
https://doi.org/10.1007/978-3-030-20890-5_18

For example, multiple shades of blue are possible for the sky, leaves of trees attain colours according to seasons, different colours of a pocket billiards (pool) ball. The aim of the colourization process is not to hallucinate the exact colour of an object (See Fig. 1), but to transfer colours plausible enough to fool the human mind. Image colourization is a widely used technique in commercial applications and a hot researched topic in the academia world due to its application in heritage preservation, image stylization and image processing.

In the last few years, Convolutional Neural Networks (CNNs) have emerged as compelling new state-of-the-art learning frameworks for Computer Vision and Pattern Recognition [1,6] applications. With the recent advent of Generative Adversarial Networks (GANs) [4], the problem of transferring colours have also been explored in the context of GANs using Deep Convolutional Neural Networks [8,13]. The proposed method involves utilizing conditional Generative Adversarial Networks (cGANs) modelled as an image-to-image translation framework.

The main contribution of this paper is proposing a variant of Conditional-GANs which tries to learn a functional mapping from input grayscale image to output colourized image by minimizing the adversarial loss, per pixel loss, classification loss and the high-level perceptual loss. The proposed model is validated using an elaborate qualitative and quantitative comparison with existing methods. A detailed ablation study which demonstrates the efficiency and potential of our model over baselines is also presented.

The rest of this paper is arranged as Sect. 2 describes previous works described in literature. Section 3 describes our proposed Perceptual-cGAN. Section 4 demonstrates a detailed qualitative and quantitative analysis of images colourized by our proposed system along with ablation studies of different objective functions. The final Sect. 5 consists of concluding remarks.

Fig. 1. The process of image colourization focuses on hallucinating realistic colours.

2 Related Work

The problem of colourization has been studied extensively by numerous researchers due to its importance in real-world applications. Methods involving user-assisted scribble inputs were the primary methods explored for the

daunting problem of assigning colours to a one-dimensional gray-scale image. Levin *et al.* [10] used user-based scribble annotations for video-colourization which propagated colours in the spatial and temporal dimensions ensuring consistency to produce colourized films. This method relied on neighbouring smoothness by optimization through a quadratic cost function. Shapiro [17] proposed a framework where the geometry and the structure of the luminance input was considered on top of user-assisted information for colourization by solving a partial differential equation. Noda et al. [14] formulates the colourization problem into a maximum a posteriori (MAP) framework using Markov Random Field (MRF) as a prior.

User-involved methods achieve satisfying results but they are severely time-consuming, needs manual labour and the efficacy is dependent on the accuracy of the user interaction. In order to alleviate these problems, researchers resorted to example-based or reference-based image colourization. Example-based methods require a reference image from the user for the transfer of colours from a source image to a target gray-scale image. Reinhard *et al.* [15] used simple statistical analysis to establish mapping functions that impose a source image's colour characteristics onto a target image. Welsh *et al.* [23] proposed to transfer only chrominance values to a target image from a source by matching the luminance and textural information. The luminance values of the target image are kept intact. The success of example-based methods rely heavily on the selection of an appropriate user-determined reference image. Hence it faces a limitation as there is no standard criterion for choosing an example image and hence depends on the user skill. Also the reference image may have different illumination conditions resulting in anomalous colour transfer.

In recent years, the use of Deep Learning methods have emerged as prominent methods for the task of Pattern Recognition, even outperforming human ability [5]. They have shown promising results in colourization by directly mapping gray-scale target images to output colors by learning to combine low level features and high-level cues. Cheng *et al.* [2] proposed the first deep learning framework using low-level features from patch, intermediate DAISY features [20] and a high-level semantic features as output to a neural network ensemble. Zhang *et al.* [25] proposed an automatic colourization process by posing it as a multinomial classification task for *ab* values. They observed that in natural images the distribution of *ab* values were desaturated and used class re-balancing for obtaining accurate chromatic components. A joint end-to-end method using two parallel CNNs and incorporation of local and global priors was proposed by Iizuka [7]. The potential of GANs in learning expressive features trained under an unsupervised scheme propelled researchers to use GANs for the task of colourization. Isola *et al.* [8] explored the idea of Conditional-GANs in an image-to-image framework by learning a loss function for mapping input to output images.

3 Proposed Method

In this section, we describe our proposed algorithm along with the architecture of the Convolutional Neural Networks employed for our method. We explore the

situation of GANs in a conditional setting [8,13] where the objective function of the generator is constrained on the conditional information and consequentially generates the output. The addition of this conditional variables enhanced the stability of the original GAN frame-work proposed by Goodfellow *et al.* [4]. Our proposed method builds upon the cGAN by incorporating adjunct losses. In addition to the adversarial loss and the per-pixel constraint, we add the perceptual loss and the classification loss in our objective function. We explain in detail about the loss functions and the networks used.

3.1 Loss Function

Conditional-GANs [8] try to learn a mapping from the input image I to output J. In this case, the generator not only aspires to win the mini-max game by fooling the discriminator but also has to be as close as possible with the ground truth. The per-pixel loss is utilized for this constraint. While the discriminator network of the cGAN remains the same when compared to the original GAN, the objective function of the Generator networks are different. The objective function of cGAN [8] is

$$\mathcal{L}_{cGAN} = min_G max_D \, \mathbb{E}_{\mathbb{I}}[\log(1 - D(G(I)))] + \mathbb{E}_{\mathbb{I},\mathbb{J}}[\log D(J)] \tag{1}$$

Here I represents the input image, J represents the coloured image, G represents the Generator and D represents the Discriminator. Isola *et al.* [8] established through their work that $\mathcal{L}_1$ is better due to their less blurring effect and also helps in reducing artifacts that are introduced by using only cGAN. Our proposed model Colourization using Perceptual Generator Adversarial Network (CuPGAN), builds up on this cGAN with incorporated additive perceptual loss and classification loss. Traditional loss functions that operate at per-pixel level are found to be limited in their attempt to capture the contextual and the perceptual features. Recent researches have demonstrated the competence of loss functions based on the difference of high-level feature in generating compelling visual performance [9]. However, in many cases they fail to preserve the low-level colour and texture information [24]. Therefore, we incorporate both the perceptual loss and the per-pixel loss for preservation of both the high-level and the low-level features. The high-level features are extracted from VGGNet model [19] trained on the ImageNet dataset [3]. Like [7], we add an additional classification loss that helps our model to guide the training of the high-level features.

For input gray-scale image I and colourized image J, the perceptual loss between I and J in this case is defined as:

$$\mathcal{L}_{per} = \frac{1}{C, H, W} \sum_{c=1}^{C} \sum_{w=1}^{W} \sum_{h=1}^{H} \| V_i(G(I^{c,w,h})) - V_i(J^{c,w,h}) \|_2 \tag{2}$$

where V_i represents a non-linear transformation by the i^{th} layer of the VGGNet, G represents the transformation function of the generator and (C, W, H) are the channels, width and height respectively. The $\mathcal{L}_{L1}$ loss between I and J is defined as:

$$\mathcal{L}_{L1} = \sum_{c=1}^{C} \sum_{w=1}^{W} \sum_{h=1}^{H} \|G(I) - \widetilde{J}\| \tag{3}$$

Here, $\widetilde{J}$ represents J in CIELAB color space. The difference between $G(I)$ and J is fundamentally the difference of their ab values.

The final objective function of the proposed generator network is

$$\mathcal{L}_{CuPGAN} = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{L1} + \lambda_2 \mathcal{L}_{class} + \lambda_3 \mathcal{L}_{per} \tag{4}$$

Here $\mathcal{L}_{adv}$ is the adversarial loss, $\mathcal{L}_{L1}$ is the content-based per-pixel loss, $\mathcal{L}_{class}$ is the classification loss and $\mathcal{L}_{per}$ is the perceptual loss and $\lambda_1, \lambda_2, \lambda_3$ are positive constants.

3.2 Generator Network

The generator network in the proposed architecture follows a similar architecture like U-Net [16]. The U-Net was originally proposed for bio-medical image segmentation. The architecture of U-Net consists of a contracting path which acts as an encoder and expanding path which acts as a decoder. The skip connections are present to avoid any information and content loss due to convolutions. The success of U-Net propelled the utilization of U-Net architecture in many works [8]. The elaborate generator network along with the feature map dimensions and the convolutional kernel sizes is shown in Fig. 2. In the encoding process, the input feature map is reduced in height and width by a factor of 2 at every level. In the decoding process, the feature maps are enlarged in height and width by a factor of 2. At every decoding level, there is a step of feature-fusion from the opposite contracting path. The final transposed convolutional layer obtained are the ab values which when concatenated with the L (luminance) channel, gives us a colourized version of the input gray-scale image. The feature vector at the point of deflection of the encoding and the decoding path is processed through two fully-connected layers to obtain the vector for the classification loss. All the layers use the ReLU activation function except for the last transpose convolutional layer which uses the hyperbolic tangent (tanh) function.

3.3 Discriminator Network

The discriminator network used for the proposed method consists of four convolutional layers with kernel size 5×5 and stride length 2×2 followed by 2 fully-connected (FC) layers. The final FC-layer uses a Sigmoid activation to classify the image as real or fake. Figure 3 shows the architecture of the discriminator network used for the proposed CuPGAN.

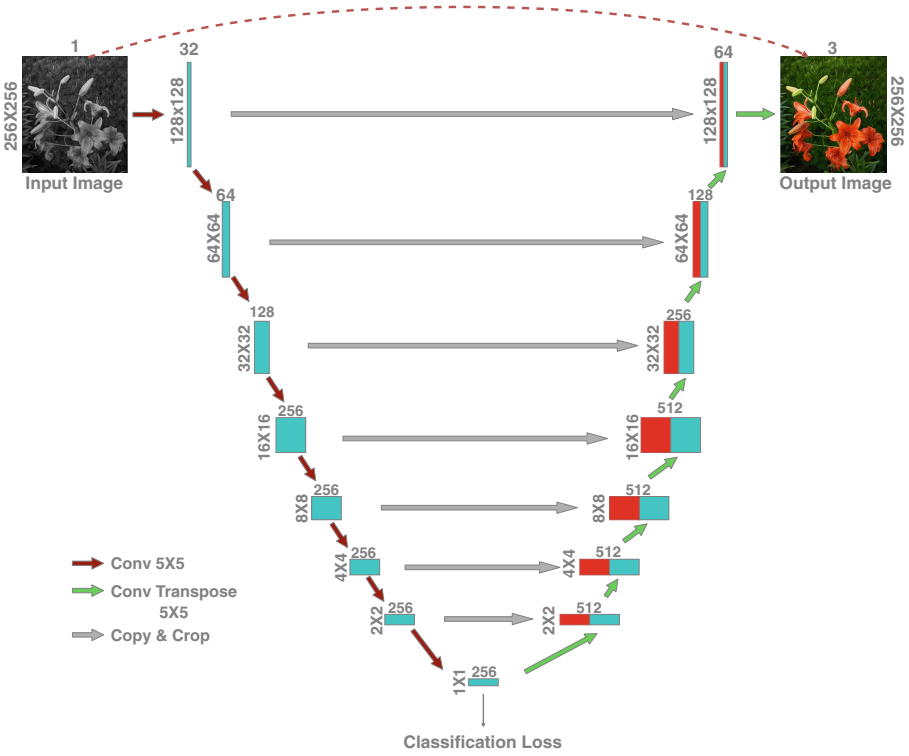

Fig. 2. Network architecture of the proposed Generator network.

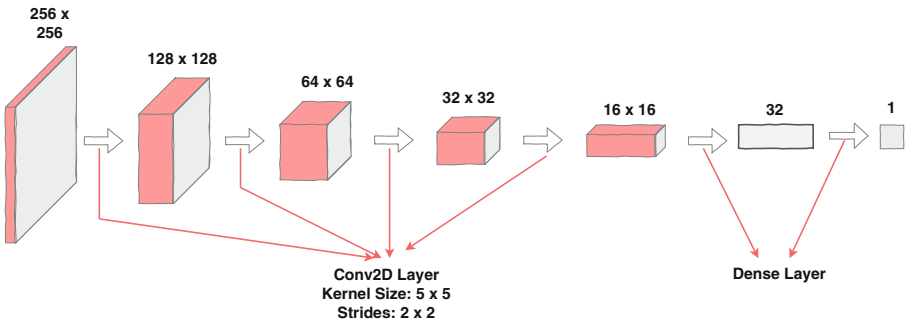

Fig. 3. Network architecture of our Discriminator Network

4 Experimental Results

4.1 Experimental Settings

All the training images are converted from RGB into the CIELAB colour space. Only the L channel of images are fed into the network and the network tries to estimate the chromatic components ab. The values of the luminance and the

chrominance components are normalized between $[-1, 1]$. The training images are re-sized to 256×256 in all cases. Every convolutional transpose layer in the decoder part of the generator network has a dropout with probability 0.5. Each layer in both the discriminator and the generator is followed by a batch-normalization layer. We use the Places365-Standard dataset [26] which contains about 1.8 million images. We filter the images that are grayscale and also those that have negligible colour variations which reduces our training dataset to 1.6 million images. For our experiments, we empirically set $\lambda_1 = 100$, $\lambda_2 = 10$ and $\lambda_3 = 1$ as the weights of the per-pixel, classification and perceptual loss respectively. We use the *relu1_2* layer of the VGGNet for computing the perceptual loss. The proposed system is implemented in Tensorflow. We train our network using the Adagrad optimizer with a learning rate of 10^{-4} and a batch-size of 16. Under these settings, the model is trained for 20 epochs on a system with NVIDIA TitanX Graphics Processing unit (GPU).

4.2 Quantitative Evaluations

We evaluate our proposed algorithm on different metrics and compare it with the existing state-of-the art colourization methods [7,25]. The metrics used for comparison are Peak Signal to Noise Ratio (PSNR) [12], Structural Similarity Index Measure (SSIM) [21], Mean Squared Error (MSE), Universal Quality Index (UQI) [22] and Visual Information Fidelity (VIF) [18]. We carry out our quantitative comparison on the test dataset of Places365. It contains multiple images of the 365 different classes used in the training dataset. The quantitative evaluations are shown in Table 1. The proposed method CuPGAN performs better in all metrics except for the VIF and UQI where it shows competitive performance on the Places365 test dataset [26].

Table 1. Comparative evaluation of PSNR, SSIM, MSE, UQI and VIF on Places365 dataset [26]

Method	Zhang [25]	Iizuka [7]	CuPGAN
PSNR	25.44	27.14	***28.41***
SSIM	0.95	0.95	***0.96***
MSE	262.08	135.89	***107.93***
UQI	0.56	***0.60***	0.58
VIF	0.886	***0.905***	0.896

4.3 Qualitative Evaluations

We demonstrate and compare our results in a qualitative manner against the existing methods [7,25]. Figure 4 displays the comparative qualitative results. We observe that in the first image Zhang *et al.* [25] hallucinates bluish colour for the grass while Iizuka *et al.* [7] produces a less saturated image compared

Fig. 4. Qualitative comparison on the test partition of Places365 dataset [26]. The results display the robustness and versatility of our proposed method for colourizing both indoor and outdoor images (Color figure online)

to ours and the ground-truth. In the second image, both methods [7,25] tend to over-saturate the colour of the structure while our result is closest to the ground-truth. In the third image of the indoor scene, Zhang *et al.* [25] tends to impart yellowish colour to the indoors contrasting our and Iizuka's [25] result and also the ground-truth. In the fourth and the last image, both methods [7,25] tend to distort colours in the grass and the road respectively. Our method does not demonstrate any such colour distortions. In the fifth image, our method tends to over-estimate the grass regions near the foothills of the mountain but displays more vibrancy as compared to [7,25]. The results of the sixth and the seventh images are satisfactory for all methods. From what we observed, Zhang *et al.* tends to over-saturate the colours of the image. The qualitative comparison ensures that not only does our method produce images of crisp quality and of adequate saturation, but it also tends to produce less colour distortions and colour anomalies.

4.4 Real World Images

In order to establish the efficiency of our model we collect some images from the internet randomly and colourize these images using all the competing methods used in the quantitative evaluations as discussed in Sect. 4.2. Figure 5 shows the colourization on real world images. We can observe that Zhang *et al.* [25] tends to over-saturate the colour components as visible from the first and the third image. Also, [25] produces some colour anomalies visible from the zebras (fourth image) and the road (fifth image). Iizuka *et al.* [7] tends to under-saturate the images as visible from the third and fourth image. The colourization from our proposed method produces more visually-appealing and sharp results comparatively.

4.5 Ablation Studies

For demonstrating the effectiveness of our loss function, we train our model using a subset of the Places365 data-set [26]. We created the subset by randomly selecting 30 different classes from the complete data-set. We train our proposed model using three different component loss functions: (1) using only $\mathcal{L}_{L1}$ loss, (2) using only $\mathcal{L}_{per}$ and (3) using both $\mathcal{L}_{L1}$ and $\mathcal{L}_{per}$ (*ours*). The use of the first loss function corresponds directly to the objective function used in the *cGAN* model proposed by Isola *et al.* [8]. We provide both qualitative and quantitative comparisons for validating our point. The qualitative results are displayed in Fig. 6. The results display the effectiveness of using both $\mathcal{L}_{L1}$ and $\mathcal{L}_{per}$ loss together over using them discretely. Table 2 displays the quantitative comparison of using the mentioned loss functions. We can infer both qualitatively and quantitatively that our objective function performs better than using only $\mathcal{L}_{L1}$ or only only $\mathcal{L}_{per}$.

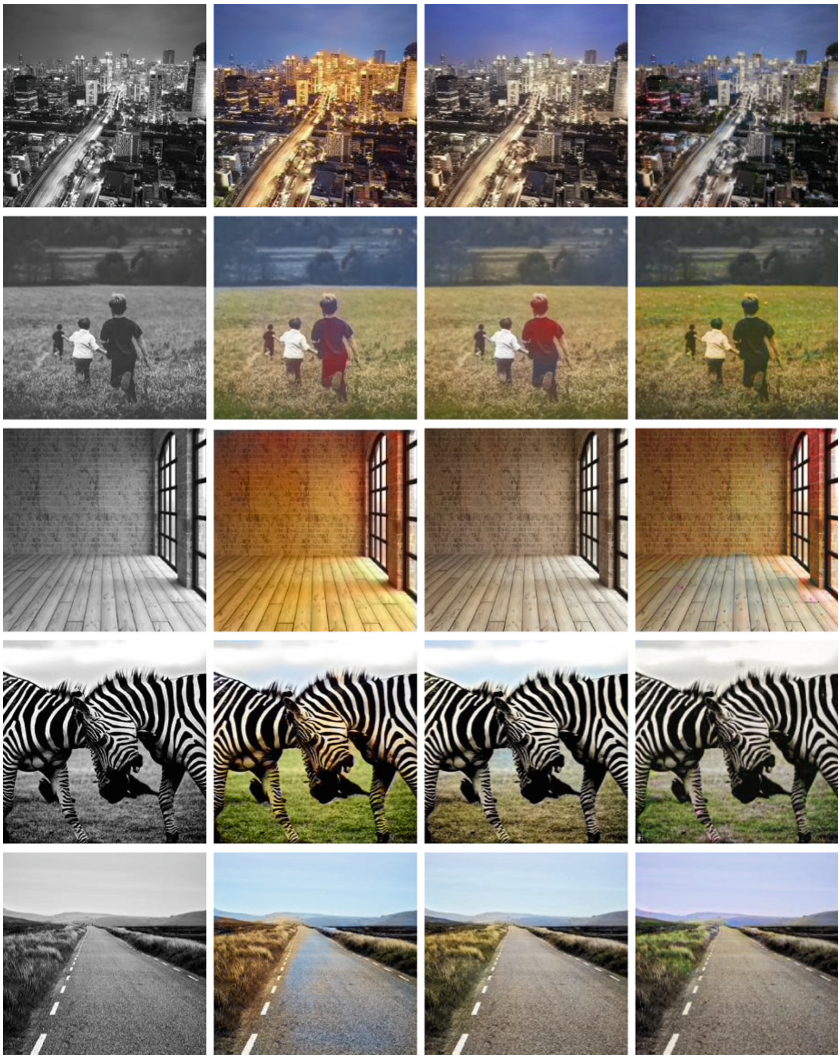

Fig. 5. Qualitative comparison on the random images from the internet

Table 2. Evaluation of PSNR, SSIM, MSE, UQI and VIF on subset test data-set from Places365 [26]

Method	$\mathcal{L}_{L1}$	$\mathcal{L}_{per}$	Ours
PSNR	22.00	17.40	***23.19***
SSIM	0.88	0.63	***0.91***
MSE	552.42	1341.66	***529.53***
UQI	***0.50***	0.20	0.49
VIF	0.85	0.53	***0.86***

Fig. 6. Ablation: Qualitative comparison on the subset-test partition of Places365 dataset [26]

4.6 Historic Black and White Images

We also test our model on 20^{th} century black and white images. Due to the type of films and cameras used in the past, we can never be perfectly sure about the type of colours and shades that was originally there. The images used have significant artifacts and have irregular borders which makes it an ill-posed task

to reckon colours. In spite of all these issues, our model is able to colourize these images producing satisfactory results. In Fig. 7 we show some examples of colourization of these historic images and we observe visually pleasing results.

Fig. 7. Colourization of historic black and white images

Fig. 8. Visual results on the CelebA dataset [11]

4.7 Adaptability

In order to establish the adaptability of our proposed method, we train our model on a very different dataset than the Places365 dataset [26]. We use the large-scale CelebFaces Attributes or the CelebA dataset [11] for training. CelebA is a large-scale dataset which contains more than 200K images with 40 attributes. The classification loss of the generator objective is calculated using the cross-entropy loss of the attributes. We provide a visual analysis and quantitative evaluation of our method on this dataset as well. We use the metrics described in Sect. 4.2. Table 3 shows the evaluation of our model on the CelebA dataset. The

Table 3. Quantitative evaluations on the CelebA dataset [11]

Metrics	CuPGAN
PSNR	**25.9**
SSIM	**0.89**
UQI	**0.70**
VQI	**0.78**

quantitative evaluations assure that our model can be employed easily to another dataset for the process of colourization. Figure 8 demonstrates the visual results on the CelebA dataset [11] supplements our claim of adaptability to versatile datasets.

5 Conclusion

In this paper, we develop on the Conditional Generative Adversarial Networks (cGANs) framework to deal with the task of image colourization by incorporating the recently flourished perceptual loss and cross-entropy classification loss. We train our proposed model CuPGAN in an end-to-end fashion. Quantitative and qualitative evaluations establish the significant enhancing effects of adding the perceptual and classification loss as compared to the vanilla Conditional-GANs. Also, experiments conducted on standard data-sets show promising results when compared to the standard exclusive state-of-the art image colourization methods evaluated using five standard image quality measures like PSNR, SSIM, MSE, UIQ and VIF. Our proposed method performs appreciably well producing clear and crisp quality colourized pictures even in cases of images picked from the internet and historic black and white images.

Acknowledgements. We would like to thank Mr. Ayan Kumar Bhunia for his guidance and help regarding implementation and development of the model. We would also like to thank Mr. Ayush Daruka and Mr. Prashant Kumar for their valuable discussions.

References

1. Bengio, Y.: Learning deep architectures for AI. Found. Trends Mach. Learn. **2**, 1–55 (2009)
2. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 415–423 (2015)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
4. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)

5. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1026–1034 (2015). https://doi.org/10.1109/ICCV.2015.123
6. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
7. Iizuka, S., Simo-Serra, E., 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. **35**(4), Article: 110, 110:1–110:11 (2016). (Proc. of SIGGRAPH 2016)
8. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. arXiv preprint (2017)
9. Johnson, J., Alahi, A., Li, F.: Perceptual losses for real-time style transfer and super-resolution. CoRR abs/1603.08155 (2016). http://arxiv.org/abs/1603.08155
10. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: ACM SIGGRAPH 2004 Papers, SIGGRAPH 2004, pp. 689–694. ACM, New York (2004). https://doi.org/10.1145/1186562.1015780. http://doi.acm.org/10.1145/1186562.1015780
11. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV) (2015)
12. Mannos, J., Sakrison, D.: The effects of a visual fidelity criterion of the encoding of images. IEEE Trans. Inf. Theor. **20**(4), 525–536 (2006). https://doi.org/10.1109/TIT.1974.1055250
13. Nazeri, K., Ng, E., Ebrahimi, M.: Image colorization using generative adversarial networks. In: Perales, F.J., Kittler, J. (eds.) AMDO 2018. LNCS, vol. 10945, pp. 85–94. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94544-6_9
14. Noda, H., Niimi, M., Korekuni, J.: Simple and efficient colorization in YCbCr color space. In: 18th International Conference on Pattern Recognition (ICPR 2006), vol. 3, pp. 685–688 (2006). https://doi.org/10.1109/ICPR.2006.1053
15. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Comput. Graph. Appl. **21**(5), 34–41 (2001). https://doi.org/10.1109/38.946629
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015). http://arxiv.org/abs/1505.04597
17. Sapiro, G.: Inpainting the colors. In: IEEE International Conference on Image Processing 2005, vol. 2, pp. II-698–701 (2005). https://doi.org/10.1109/ICIP.2005.1530151
18. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006). https://doi.org/10.1109/TIP.2005.859378
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)
20. Tola, E., Lepetit, V., Fua, P.: Daisy: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5), 815–830 (2010)
21. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
22. Wang, Z., Bovik, A.: A universal image quality index. IEEE Signal Process. Lett. **9**(3), 81–84 (2002). https://doi.org/10.1109/97.995823

23. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2002, pp. 277–280. ACM, New York (2002). https://doi.org/10.1145/566570.566576. http://doi.acm.org/10.1145/566570.566576
24. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. arXiv preprint arXiv:1701.05957 (2017)
25. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 649–666. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_40
26. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: a 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(6), 1452–1464 (2018)