GAN-Based Image Restoration and Colorization

Aliyah Kabeer, Manali Tanna, K. N. Milinda, Mohammed Uzair Rizwan, and Pooja Agarwal

1 Introduction

Over the past two decades, mobile phone imagery has seen a significant evolution. Thicker lenses or huge image sensors cannot be put on cell phones due to the structure's restriction on the size of the camera module. Both the image's volume and picture quality are impacted by this. It is challenging to capture perceptually pleasing images in light and dark environments especially during rapid movement (Gampala et al. 2020). The best images frequently require extensive postprocessing on the computer. The aesthetic enhancement of a photograph can be achieved through image restoration and image colorization. In order to make a grayscale image more aesthetically pleasing and perceptually relevant, a procedure known as image colorization is used (Sophia et al. 2022). It is acknowledged as a challenging task that frequently necessitates familiarity with the image content beforehand and manual changes to obtain artifact-free quality. There is no right solution to this task because objects might have multiple colors and there are numerous ways to assign colors to pixels in an image. On the other hand, image restoration entails recovering the original image from its noisy and obscured counterpart. It focuses on the elimination or mitigation

A. Kabeer (✉) · M. Tanna · K. N. Milinda · M. U. Rizwan · P. Agarwal
Department of Computer Science and Engineering, PES University, Bangalore, India
e-mail: aliyahk888@gmail.com

M. Tanna
e-mail: manalitanna29@gmail.com

K. N. Milinda
e-mail: milimilindakn@gmail.com

M. U. Rizwan
e-mail: uzairrizwan15@gmail.com

P. Agarwal
e-mail: poojaagarwal@pes.edu

N. R. Shetty et al. (eds.), *Advances in Communication and Applications*, Lecture Notes in Electrical Engineering 1105, https://doi.org/10.1007/978-981-99-7633-1_40

of noise degradation brought on by camera motion blur or misfocus (Nguyen et al. 1604).

Image colorization attempts in the past often incorporated user input. Researchers switched to color transfer, which was based on the resemblance between a provided reference image and the input, to lessen the user workload. Deep learning-based automatic colorization techniques were suggested as a way to further lessen this dependence. These techniques identified the connection between the colorful image from a dataset and the monochrome input. Deep learning-based restoration techniques replaced manual deblurring attempts using editing tools. Autoencoders were frequently utilized for image restoration tasks as they included an encoder–decoder framework and symmetric convolutional–deconvolutional layers to extract the key aspects of the image and ignore the degradations while recovering it. However, substantial advancements have been made in these two tasks individually owing to the development of generative adversarial networks (GANs) (Generative adversarial networks 2020). Using an adversarial network, GANs present a generative model that can produce high-quality natural photos that gradually develop to generate more and more realistic-looking data. In addition to being able to produce extremely high-quality synthetic data, this system may also be used to generate images, colorize, and restore them. The majority of current methods for image restoration and colorization use GANs and their variations to create realistic images.

The proposed GAN aims to utilize loss functions specific to both restoration and colorization and uses an architecture that encompasses the features of both. This explores a new side of GANs and introduces a novel architecture for image enhancement tasks. It also paves the way for a new avenue of research that can exploit GANs where two or more tasks can be achieved together. GAN's are usually used to perform one single task at a time and previous research highlights this point as the GAN models proposed earlier in the image enhancement domain only either restore or colorize, but don't perform both together within the same architecture.

2 Related Work

In this section, the various advancements made in the field of image restoration and colorization are discussed. Traditional methods that aimed to tackle this problem involved some degree of manual intervention or effort for picking the right colors and deblurring and sharpening images using image editing software. With the advent of deep learning techniques, these tasks were automated by training models to learn and generalize color features and restoration functions from a selected distribution of input–output paired images. With the advent of GANs, this was further automated by eliminating the need of a human to evaluate the loss function. Instead, the GAN architecture uses two components—the generator and the discriminator that work together to automatically ensure that the loss function is performing well to produce very realistic images.

For image restoration tasks, Wang et al. (2021) focus on restoring facial images only and apply restoration techniques to extract high-quality images from the corresponding low-quality image counterpart. They attempt to use pretrained face GANs as generative facial priors (GFP) to successfully retain image information and maintain quality. However, the model could not perform well when the degradations were severe or the facial images were of varied poses. One of the main reasons for this was the fact that the model was trained on synthetically degraded data instead of real-world degraded images. Tomosada et al. (2021) propose an image deblurring method using GANs that requires less computational complexity to train and makes use of discrete cosine transform (DCT). The model is trained on a customized dataset including images with large motion blurs and introduces a novel loss function that compares images in the frequency domain. The authors of Pan et al. (2021) utilized image priors from a GAN that was trained on a large dataset. Deep generative priors were employed to replace the missing semantics in the images. Scaling the training images and training an encoder that maps the target image to the latent code helped the architecture's overall performance. This made accurate image reconstruction possible, but it also presupposed known degradation. The model does not generalize well to unknown degradation and performed poorly on real-world blind restoration. The authors in Ledig et al. (2017) introduce a content loss driven by perceptual similarity rather than similarity in pixel space to restore the features of degraded images while training. The deep residual network of the GAN can successfully recover photo-realistic textures from heavily down-sampled images.

Image colorization is an unconstrained problem with no unique solution, and it has been studied as a challenging problem for over many decades. The authors in Thomas et al. (2018) develop a model to efficiently colorize black and white images and videos, and eliminate issues such as—inconsistent colorization within individual objects, undersaturation, and oversaturation of images. The model was developed on the premise that: neighboring image frames with neighboring pixels that have similar intensities should ultimately have similar colors. The authors used a long short-term memory (LSTM) for dealing with grayscale videos and convolutional neural networks (CNNs) for images. The authors in Sarmai and Sadhu (2022) compare the performance of convolutional autoencoders and GANs for the colorization of grayscale images. They observed that the images colorized by the autoencoders lost their sharpness and were blurry in nature, whereas those colorized by the GAN are nearly identical to their ground truth counterparts. This further emphasizes the choice of using GANs in the proposed model. Nazeri et al. (2018) proposed a GAN architecture that included one-sided label smoothing, batch normalization, and RELU activation function. Instead of the RBG color space, they used the CIE Lab color space, where L is the lightness channel and a and b are the color channels. They observed that this prevented any sudden variations in both color and brightness as in RGB. Xiao et al. (2019) were inspired by Cycle-GAN (Zhu et al. 2017) and the concept of image-to-image translation. Their method uses unpaired images for training and direct prediction of color in the RGB color space, which eases the process of training data collection. However, their model was only able to generate reasonable color instead of restoring original color for input grayscale images. The authors of

Mourchid et al. (2020) propose an additional discriminator in their GAN which works in the feature domain. Such an architecture is proposed so that the generator learns to produce high-frequency features instead of noise from the normal distribution. The model colorizes images realistically and requires less artifacts. As future work, they suggested a typical Resnet architecture to retain low-scale features making it shallow. Sankar et al. (2020) study the use of GANs for colorization using the perceptual loss function. The model was trained on the open source CIFAR-100 dataset. Additionally, they did a comparative study on the same dataset using three different models—the U-Net CNN, GAN, and GAN with perceptual loss. After evaluating the results and comparing the three models, the authors' concluded that the GAN with perceptual loss gave the best results for colorization. This further strengthens the choice of using the same loss function as an integral part of the proposed GAN architecture.

Therefore, in their own separate domains, brilliant research has been carried out in restoration as well as colorization of images. As a joint task very few researchers have been able to provide promising results. Compared with early deep learning methods like CNN and LSTM, autoencoders have outperformed in the tasks of image restoration. With the advent of GANs or generative adversarial networks, they have outperformed even autoencoders in both restoration and colorization. GANs can produce more realistic and better-quality images. Although computation in these methods is expensive, overhead is reduced as the loss function is learnt automatically. To conclude this section, it is identified that GAN-based image restoration and colorization has a wide scope, and can contribute significantly to this domain.

3 Proposed Methodology

Most existing systems focus solely on either restoration or colorization of images or on both sequentially. Moreover, limitations on colorization such as over/undersaturation and the inability to capture the different colors of intricate images have been a common recurring problem. The reason why previous algorithms and models for colorization didn't work remarkably well is because there was still a human involved in hand-coding a key step, which is evaluating whether or not the restored or colorized image "looked good." So over the last couple of years, generative adversarial networks (GANs) have increasingly been replacing vanilla U-Net architectures and autoencoders (Sarmai and Sadhu 2022) in order to solve this problem, as it has proven to effectively learn the loss function with the help of its discriminator that can evaluate the generated image and also ensure more realistic feature generation. The proposed system aims to utilize this advantage of GANs in order to both effectively restore and colorize degraded grayscale images within the same architecture. The proposed methodology includes the details of the GAN architecture, the preprocessing stage of the model, followed by the breakdown of the GAN into its two different components—the generator and the discriminator.

The foundation of GANs is built on a generator that learns to generate new images and a discriminator that learns to differentiate between artificial and original images.

A conditional setup is used in GANs, which ensures that the generator and discriminator are both dependent on extra data (such as class labels or information from other modalities). The ideal model is thus capable of learning multi-modal mapping from inputs to outputs by being fed various contextual data. This conditional setup is what is known as a conditional GAN (Zhao et al. 2019). The proposed algorithm follows that of the conditional GAN algorithm and involves two separate modules—the generator and the discriminator, that are trained simultaneously.

3.1 Preprocessing

As part of the preprocessing stage, the images from the dataset are first resized, following which they are converted from the RGB color space to the L^*a^*b color space (Nazeri et al. 2018). The L channel of the degraded image serves as the grayscale image and its L^*a^*b counterpart as the colored image. The images are then split into the train and test datasets before being fed into the generator.

3.2 Generator Module

The generator takes a degraded grayscale image as input and generates a restored, colorized version of the same. The generator consists of a U-Net architecture (Ronneberger et al. 2015) as seen in Fig. 1, which includes a series of encoders and decoders. In order to speed up the training, the proposed model utilizes the pretrained ResNet50 model (He et al. 2016) in the first few layers of the encoder. This ensures the model can identify features faster and take significantly less time for training. This is followed by six decoder blocks that upscale the image to give a three-channeled output or a colored image.

3.3 Discriminator Module

The discriminator primarily serves as a binary classifier and will classify the input image as being either real or fake. It consists of a series of convolutional layers as seen in Fig. 2 with "Relu" activation function, followed by max pooling layers. The final layer utilizes the sigmoid activation function in order to generate a value between 0 and 1, indicating whether the input image is classified as fake or real, respectively.

The final part of the model includes the loss functions used in the training of the generator and discriminator. The first loss function is the GAN loss or adversarial loss. This minimax loss is traditionally used to train the generator and discriminator in sync with one another. The discriminator uses the cross-entropy loss function for training. The generator utilizes the VGG-19 pretrained model (Simonyan and

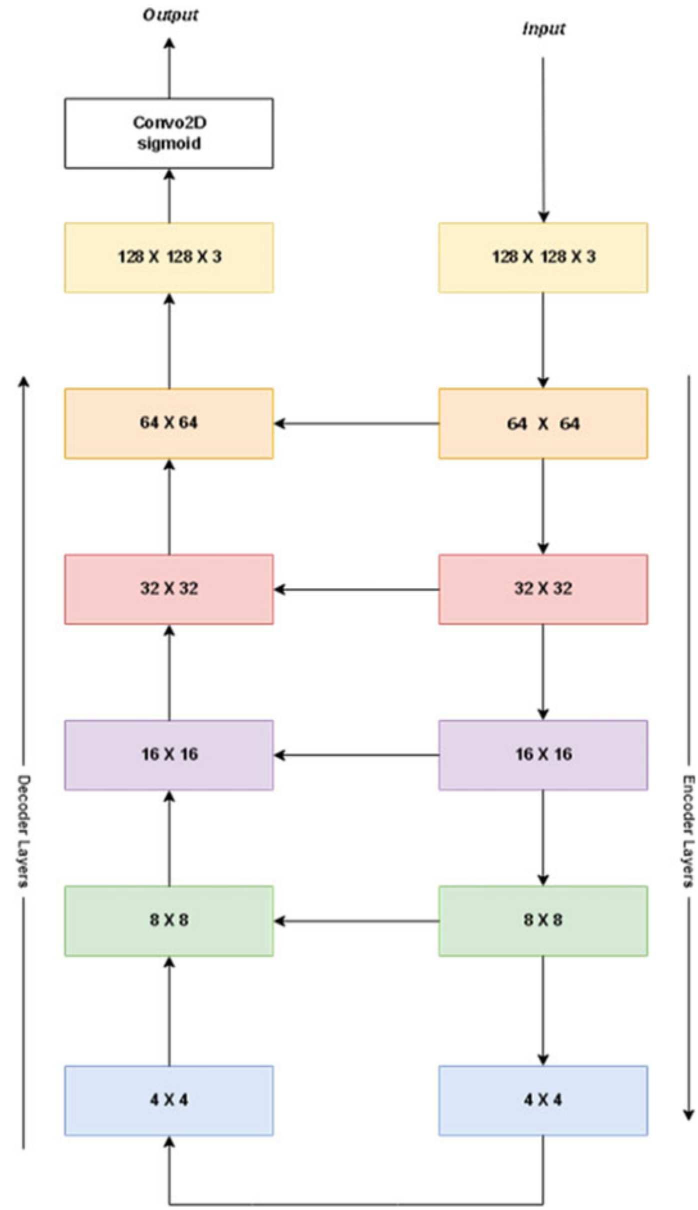


Fig. 1 Proposed generator architecture

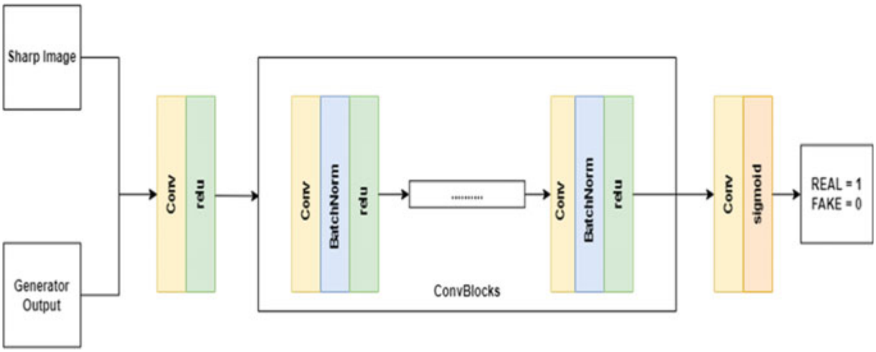


Fig. 2 Proposed discriminator architecture

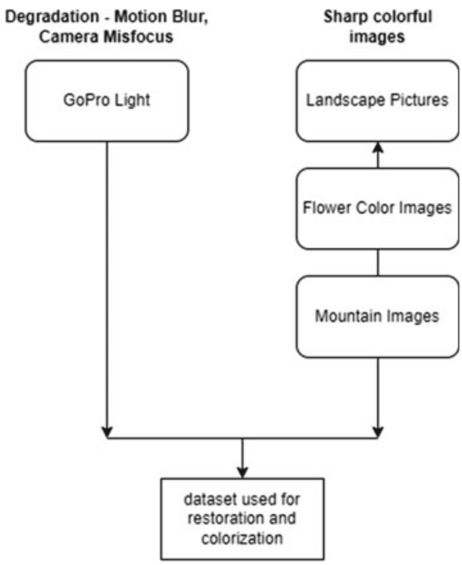
Zisserman 2015) to obtain the feature-level representations of the generated image and the original image and performs a feature-wise comparison of the two. This loss function is known as the perceptual loss function and is used in combination with the L1 loss and the adversarial loss to update the generator's weights. The Adam optimizer is utilized with a learning rate of 0.005 for both the generator and the discriminator. Because the generator model and discriminator model are trained simultaneously in a zero-sum game, they are challenging to train. In other words, while one model improves, the other model suffers. Finding a point of equilibrium between the two conflicting concerns is the aim of training two models. It also implies that the nature of the optimization problem being tackled changes each time the parameters of one of the models are modified. As a result, a dynamic system is created. The technical difficulty of simultaneously training two competing neural networks is that they might not converge. Plotting the generator and discriminator's losses provides insight into the model's convergence.

4 Experimental Setup and Results

4.1 Dataset

The dataset required for the project includes images containing different classes of objects pertaining to nature so that the model is versatile and able to restore and colorize all kinds of scenery images. Colored, sharp images are expected from the dataset as they will be used as the ground truth images while training the model. Since GANs require huge amounts of data to be trained well, a relatively large and vast dataset is required. Previous work in this field implied that working with datasets having too many different classes might result in poor performance of the model. Hence, a dataset with a limited but varied number of classes is preferred, and for this reason a dataset of 5500 images is curated, consisting of images of colorful

Fig. 3 Curation of dataset



scenery (Arnaud 2020), flowers (Belitskaya 2017) and mountains (https://www.kaggle.com/puneet6060/intel-image-classification) to focus on the colorization aspect of the model training. This was combined with the GoPro Light dataset (Tomosada et al. 2021) which consists of 2103 blurred and sharp image pairs for training. The images are captured by a high-speed camera. This allows the model to learn how to restore naturally degraded images affected by blur or motion blur.

The images of the newly curated dataset consists of 6840 training images and 763 testing images. They are resized to 128 $\times$ 128 and converted to grayscale as a part of the training phase of the model. As seen in Fig. 3, the final dataset used in training the model comprises of a combination of the GoPro Light dataset which consists of blur images and their clear counterparts, along with the manually curated dataset consisting of only nature-specific images.

4.2 Implementation

The proposed GAN architecture is implemented using tensorflow, keras. In total three models are implemented. The first model COLOR-GAN is the proposed architecture trained on just the colorization dataset and it performs no restoration. The input is a sharp grayscale image and the output is the colorized output. If the input is blurry then it doesn't perform any restoration on it and the colorization is not up to the mark. The second model RESTORE-GAN, uses the architecture trained only on the GoPro dataset and only performs restoration. The input is a blurry image and the output is the corresponding sharp image. This model performs well in terms of restoration, no matter the color of the input. Finally, the third model restore and colorize images GAN (RCI-GAN) uses the proposed architecture on the curated dataset. It takes as input a degraded (blurry) grayscale image and outputs the restored and appropriately colored image. Each of the models was trained for 100 epochs and tested on the test images from the GoPro light dataset (Tomosada et al. 2021). Figure 4 depicts the results of the COLOR-GAN, Fig. 5 depicts the results of RESTORE-GAN, and Fig. 6 depicts the results of RCI-GAN, respectively. The three models are evaluated based on image quality-specific metrics like peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). The PSNR metric displays the relationship between a signal's maximum achievable power and the power of corrupting noise that compromises the accuracy of its representation. PSNR, a variant of MSE, focuses on pixel-by-pixel comparison. SSIM is related to the quality and perception of the human visual system (HVS color model). The SSIM represents picture distortion as a combination of three elements, namely loss of correlation, luminance distortion, and contrast distortion, as opposed to utilizing conventional error summation techniques. The PSNR and SSIM metrics for all three models are recorded on the test data and tabulated in Table 1.

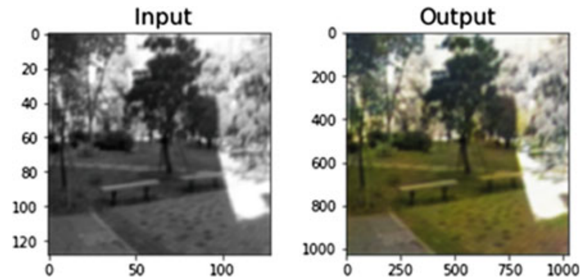


Fig. 4 COLOR-GAN output

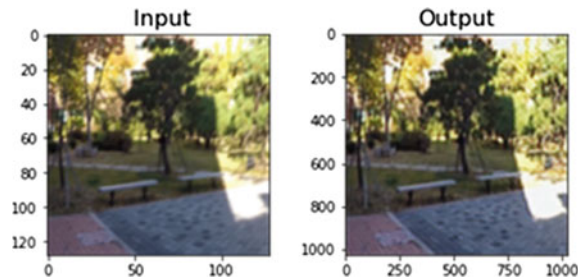


Fig. 5 RESTORE-GAN output

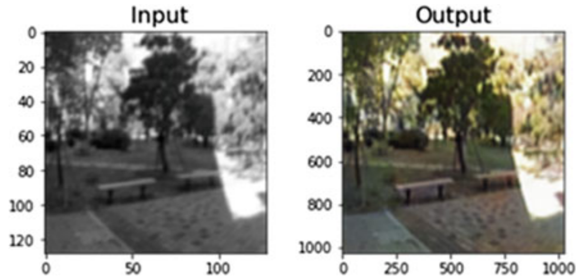


Fig. 6 RCI-GAN output

Table 1 Comparison of proposed models

	PSNR	SSIM
COLOR-GAN	22.96	0.74
RESTORE-GAN	25.64	0.93
RCI-GAN	23.22	0.80

4.3 Results

The proposed RCI-GAN trained on the curated dataset is able to effectively remove motion blur and colorize both degraded as well as sharp grayscale images specific to the scenery class of images. As the model has been trained on a dataset containing images of nature and scenery, the model is fine-tuned to such images only at present. The model assigns reasonable color to the different features of the image, and is able to produce perceptually meaningful results. In terms of image quality and restoration, the model gives an average PSNR value of 23 and an average SSIM value of 0.8 when tested on real-world sample images. Figure 7 depicts the outputs obtained on testing the model on originally grayscale real-world samples. Figure 8 shows the outputs obtained on colored real-world samples that were converted to grayscale as a part of preprocessing, and Fig. 9 depicts the model's performance on predominantly green landscape images that were also converted to grayscale before being fed into the model.

4.4 Comparison of COLOR-GAN, RESTORE-GAN, and RCI-GAN Results

From Table 1 it can be concluded that when trained on the specific dataset, the proposed GAN architecture is able to achieve a good trade-off between image restoration and image colorization with acceptable evaluation metrics. The PSNR and SSIM values obtained for the RCI-GAN lie in between those obtained from COLOR-GAN and RESTORE-GAN, thereby proving that the proposed model can

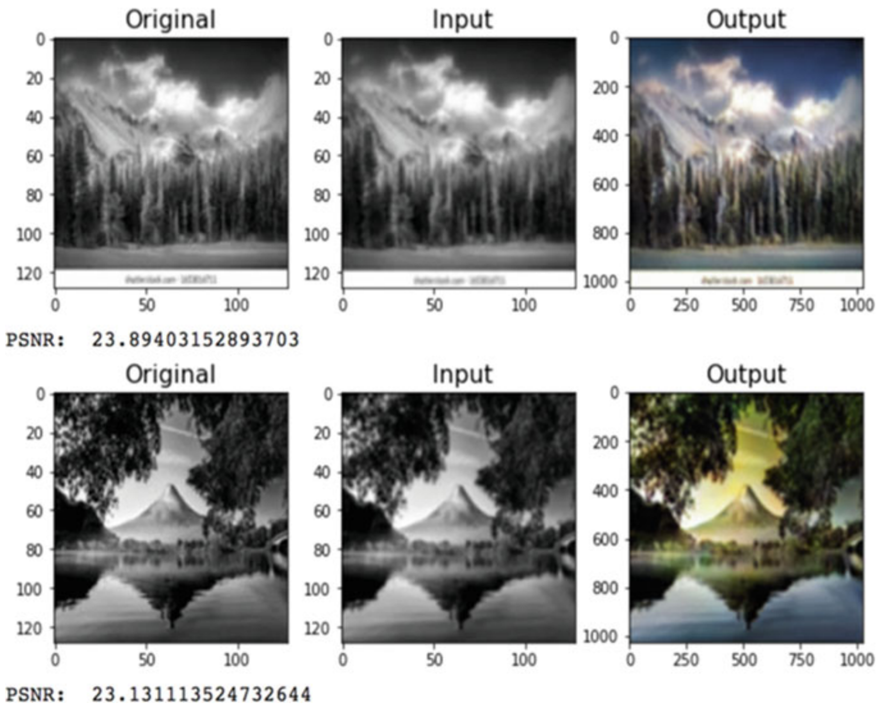


Fig. 7 Visual results on real-world data

successfully combine both tasks within the same architecture without significant loss in performance.

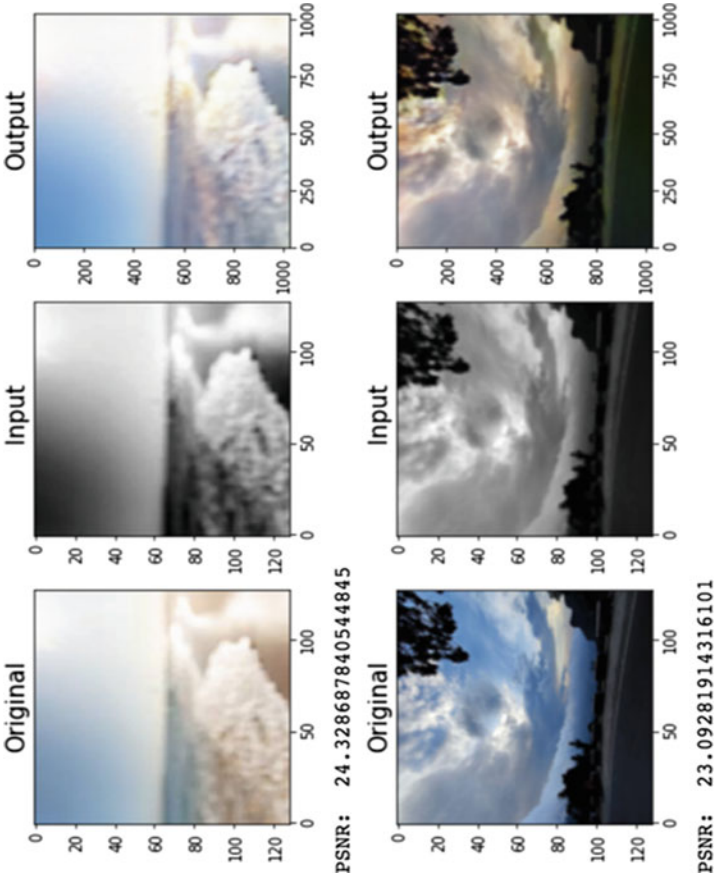


Fig. 8 Visual results on real-world data

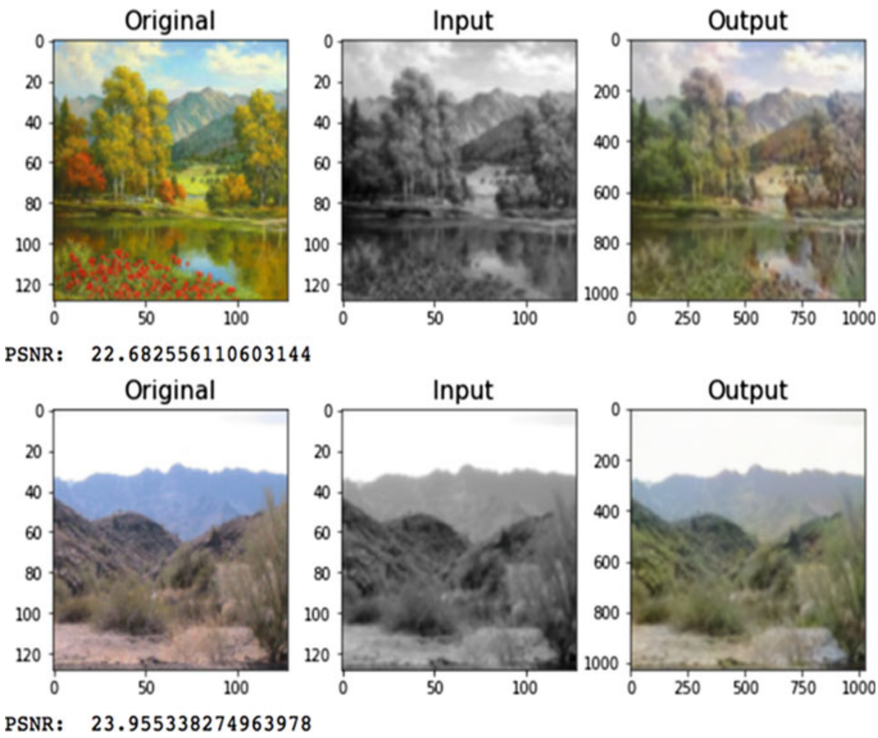


Fig. 9 Visual results on real-world data

4.5 Comparison of RCI-GAN Results with Existing Models

Evaluation metrics of the proposed model are compared with existing GAN models as stated in published research works. GFP GAN (Wang et al. 2021) performs blind face restoration and colorization, but is specific only to facial images. DeblurDCT GAN (Tomosada et al. 2021) and ESRGAN (Wang et al. 2018) perform deblurring and restoration of naturally blurred images but do not perform colorization. There are no models that do exactly what the proposed GAN offers and Table 2 provides a good comparison to evaluate results. It compares the PSNR and SSIM values of the proposed GAN with the existing models, performing either image restoration, colorization, or both. The model is successfully able to perform two tasks: image restoration and image colorization within a single GAN architecture, thereby reducing model complexity and eliminating the need for a second GAN. The RCI-GAN compares well with the evaluation metrics of existing models as depicted clearly in Table 2.

Table 2 Comparison of RCI-GAN with existing models

	PSNR	SSIM
RCI-GAN	23	0.8
GFP GAN	25	0.6
Deblur GAN	29	0.9
ESRGAN	23	0.6

5 Conclusion

While GAN architectures can be utilized in a wide variety of applications, it's rare to find a GAN architecture that can perform multiple tasks at once. This paper introduces a new avenue of research where a single GAN model can perform two tasks at the same time: image restoration and image colorization. The proposed restore and colorize images GAN (RCI-GAN) effectively restores and colorizes nature-specific images within a single GAN architecture. This eliminates the need for utilizing two separate GANs for both tasks individually, thereby presenting a novel architecture for training the GAN to focus on both tasks simultaneously. The model utilizes a customized dataset (GoPro, landscape, flowers, and mountains) and loss functions (L1 loss, adversarial loss, and perceptual loss) specific to both restoration and colorization to construct an architecture that encompasses the features of both image enhancement tasks. The image quality of the restored and colorized images have been calculated and compared with the existing models using metrics like peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) and shows promising results.

6 Limitations and Future Work

Figure 7 shows an image depicting the reflection of a mountain on a water body, surrounded by trees on the top left and right corners of the image. The model assigns the color green to the top part of the image and blue to the bottom part of the image. This is how the model interprets the image. However, the interpretation of the image could vary. Hence, one of the limitations include accurately colorizing the image according to human perception. Moreover, when the degradation is severe or the image sample is out of distribution, the model produces unsatisfactory results which could be improved by training the GAN on a larger class of images and for higher number of epochs. Loss graphs could be plotted at each instance of training to analyze and optimize convergence. Since the model produces good results on scenery/nature-specific images and generalizes well for real-world samples, it shows good potential for the model to be extended to all classes of images. It also paves the way for a unique area of research that can exploit GANs where two or more tasks can be achieved together.

References

Arnaud R (2020) Landscape pictures, version 1. https://www.kaggle.com/datasets/arnaud58/landscape-pictures. Accessed 2022

Belitskaya O (2017) The dataset of flower images, version 7. https://www.kaggle.com/code/olgabelitskaya/the-dataset-of-flower-images/data. Accessed 2022

Gampala V, Sunil Kumar M, Sushama C, Fantin Irudaya Raj E (2020) Deep learning based image processing approaches for image deblurring. Mater Today Proc. https://doi.org/10.1016/j.matpr.2020.11.076

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Commun ACM 63(11):139–144. https://doi.org/10.1145/3422622

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of IEEE conference computer vision pattern recognition (CVPR), pp 770–778

Intel image classification challenge, mountain class. https://www.kaggle.com/puneet6060/intel-image-classification

Ledig C, Theis L, Huszar F, Caballero J, Cunning-ham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Twitter WS (2017) Photo-realistic single image super-resolution using a generative adversarial network

Mourchid Y, Donias M, Berthoumieu Y (2021) Automatic image colorization based on multi-discriminators generative adversarial networks. In: 2020 28th European signal processing conference (EUSIPCO), pp 1532–1536. https://doi.org/10.23919/Eusipco47968.2020.9287792

Nazeri K, Ng E, Ebrahimi M (2018) Image colorization using generative adversarial networks. In: Perales F, Kittler J (eds) Articulated motion and deformable objects. AMDO 2018. Lecture notes in computer science, vol 10945. Springer, Cham. https://doi.org/10.1007/978-3-319-94544-6 9

Nguyen T, Mori K, Thawonmas R (2016) Image colorization using a deep convolutional neural network. arXiv preprint arXiv:1604.07904

Pan X, Zhan X, Dai B, Lin D, Loy CC, Luo P (2021) Exploiting deep generative prior for versatile image restoration and manipulation. In: IEEE transactions on pattern analysis and machine intelligence

Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham

Sankar R, Nair A, Abhinav P, Mothukuri SKP, Koolagudi SG (2020) Image colorization using GANs and perceptual loss. In: 2020 international conference on artificial intelligence and signal processing (AISP), pp 1–4. https://doi.org/10.1109/AISP48273.2020.9073284

Sarmai A, Sadhu A (2022) Image colorization using convolutional autoencoders and generative adversarial networks, vol 9. IARJSET, pp 86–99. https://doi.org/10.17148/IARJSET.2022.9115

Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: ICLR

Sophia SGG, Umar S, Visvas N (2022) An efficient method for blind image restoration using GAN. In: 2022 international conference on innovative computing, intelligent communication and smart electrical systems (ICSES), pp 1–8. https://doi.org/10.1109/ICSES55317.2022.9914234

Thomas BS, Dogra R, Dixit B, Raut A (2018) Automatic image and video colourisation using deep learning. In: 2018 international conference on smart city and emerging technology (ICSCET), pp 1–4. https://doi.org/10.1109/IC-SCET.2018.8537308

Tomosada H, Kudo T, Fujisawa T, Ikehara M (2021) GAN-based image deblurring using DCT loss with customized datasets. IEEE Access 9:135224–135233. https://doi.org/10.1109/ACCESS.2021.3116194

Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Loy CC, Qiao Y, Tang X (2018) ESRGAN: enhanced super-resolution generative adversarial networks. ECCV Workshops

Wang X, Li Y, Zhang H, Shan Y (2021) Towards real-world blind face restoration with generative facial prior. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 9164–9174. https://doi.org/10.1109/CVPR46437.2021.00905

Xiao Y, Jiang A, Liu C, Wang M (2019) Single image colorization via modified cyclegan. In: 2019 IEEE international conference on image processing (ICIP), pp 3247–3251. https://doi.org/10.1109/ICIP.2019.8803677

Zhao JLL, Bai H, Zeng AWB, Zhao Y (2019) Simultaneous colordepth super-resolution with conditional generative adversarial networks. arXiv:1708.09105

Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2223–2232