# Deep-Learning Based Colorization of Grayscale Images and Videos

**Ramsagar Shreya[1], Afsha Shareefa[2], Mustafa Ali Khan[3], Mr. R. Vinod Kumar[4]**

*Student, Department of Artificial Intelligence & Data Science, Methodist College of Engineering & Technology, Hyderabad, Telangana, India [1 2 3]*

*Assistant Professor, Department of Computer Science Engineering, Methodist College of Engineering & Technology, Hyderabad, Telangana, India [4]*

**Abstract–** The area of automatic colorization of grayscale images and videos is an important area of research that has attracted a lot of attention in recent times because of the advancements in deep learning techniques. The paper proposes a unified framework for colorizing images and videos based on the implementation of a Conditional Generative Adversarial Network and a U-Net-based generator. The colorization of images and videos is performed in the Lab* color space, where the luminance channel is used to colorize the output.

The proposed method differs from other existing colorization techniques in that it uses a single model to process both images and videos, whereas other techniques require different models to process these two different types of inputs. The colorization of the video is performed by processing each frame of the video separately and then reconstructing the output video by arranging the processed frames in sequence. The proposed method does not require any reference images and can be performed efficiently.

**Keywords:** Image Colorization, Video Colorization, cGAN, U-Net, Deep Learning, Computer Vision, Lab Color Space

## 1. INTRODUCTION

The process of converting the grayscale images and videos into realistic colors is an important area of research in the field of computer vision. The process of colorization of images as well as videos is considered an ill-posed problem, as there are various possibilities for the colorization process. The traditional methods for the process of colorization involved high amounts of human intervention, which is considered time-consuming.

With the introduction of deep learning techniques in the field of computer vision, the process of colorization of images as well as videos is significantly improved. The Generative Adversarial Networks are found to be highly efficient for the process of image-to-image translation. The complex data distribution of the images can be learned with the help of the Generative Adversarial Networks, as the process of colorization is possible with such networks.

In this work, a unified approach based on deep learning has been proposed for colorization, applicable to both images and videos. The approach is based on a Conditional Generative Adversarial Network, denoted as cGAN, and includes a U-Net model. In the approach, the luminance channel, denoted as L, of the input grayscale image is considered, and chrominance channels a and b are generated. Unlike other approaches, a single model is utilized for colorization, applicable to both images and videos. In the approach, a different model is not utilized for colorization of images and videos.

The approach for colorization of videos, as proposed in this work, is based on a frame-based approach. In this approach, each frame of the input video is considered as an image, and colorization is performed using a cGAN model. After colorization, the frames are combined to obtain the output video. The approach has been developed to provide an automated, scalable, and user-friendly approach for colorization of grayscale images and videos, considering a single model for colorization, applicable to both images and videos, and avoiding the need to annotate color data.

## 2. LITERATURE SURVEY

A detailed survey and benchmark on the techniques used for image colorization were presented by Anwar et al. in their research paper (2025). It describes different techniques using deep learning methods and their advantages and disadvantages. Even though this paper provides a detailed overview, the unified system is still missing in the real-time environment, which is addressed in this work using the end-to-end model [1].

In their research paper (2024), Reddy et al. proposed advanced adversarial techniques for grayscale video colorization using deep learning techniques. Their system provides quality and consistency in the image colorization process. However, the techniques proposed by the authors involve high computational complexity, which is addressed in this work using the optimized processing techniques [2].

Tadiparthi et al. proposed a grayscale image colorization model using the CNN model with PyTorch in their research paper (2024). Their model provides quality in the feature extraction process. Realistic colors are still missing in the model, which is addressed in this work using the GAN-based model [3].

Abhiram et al. (2023) discussed parallel computing and deep learning techniques for colorization of visual media. The proposed method is efficient in terms of processing time. However, it has made the system more complex. Our approach is a balance of efficiency and simplicity. An optimized architecture has been used in the proposed approach [4].

Chatterjee & Iyer (2022) discussed a GAN-based approach for the colorization of grayscale images. The proposed approach is efficient in terms of accuracy. However, transfer learning is not more adaptive in terms of changes in input images. Our approach is more adaptive in terms of changes in input images because it uses direct learning from data [5].

Singh et al. (2023) discussed a deep learning-based approach for the automatic colorization of videos considering consistency. The proposed approach is efficient in terms of reduction of flickers. However, it has made the system more computationally expensive. Our approach maintains consistency in a computationally efficient manner [6].

A hybrid CNN-GAN approach was proposed by Lee et al. (2024) to enhance the quality of the images during the colorization process. Even though the system enhances the quality of the images, the system is limited in terms of the training process. The system is modified to enhance the quality of the images by incorporating the hybrid loss function [7].

Mehta et al. (2024) proposed various advanced deep learning techniques to enhance the quality of the images during the colorization process. The system is modified to enhance the quality of the images by improving the feature learning process, thus ensuring robustness [8].

Nguyen et al. (2025) proposed a hybrid approach based on the GAN algorithm to enhance the quality of the images during the colorization process. Even though the system is effective, the system is limited in terms of efficiency. The system is modified to enhance the quality of the images by improving the efficiency [9].

Zhao et al. (2023) introduced the Vision Transformer-based approach for the automatic colorization of images. The approach is capable of efficiently handling the global context information; however, the approach requires high memory and computational resources. The proposed U-Net and GAN-based approach is an efficient solution compared to the existing approaches [10].

In the research article by Patel & Verma (2022), the authors have introduced the review of the various techniques used for the colorization of images and videos with the help of deep learning approaches. Even though the authors have discussed various techniques, the implementation is lacking. The current work is an addition to the complete solution [11].

Gupta et al. (2024) introduced an efficient approach for the color restoration of images with the help of the GAN and U-Net architectures. The approach faces difficulties while handling the inputs; the current work is an enhancement to the existing approach in terms of the adaptability [12].

## 3. PROPOSED WORK

The proposed system offers a unified framework based on deep learning techniques for the automatic colorization of images as well as videos. Unlike other techniques, which use different techniques for different types of data, the proposed system offers a unified framework using a conditioned Generative Adversarial Network (cGAN) for efficient image as well as video colorization.

The key component of the system is a generator network based on a U-Net architecture and conditioned GAN. The system utilizes the Lab color space, where the luminance component (L) of the grayscale image is taken as input, and the chrominance components (a and b) are predicted. This is because the luminance component of images is more significant in preserving structural information than the chrominance component.

In the case of image colorization, the input image will undergo a series of preprocessing steps, where the image will be resized and then converted to Lab color space. The L channel of the image will be passed through the generator network, and then the a and b channels will be generated.

In the case of video colorization, a different approach will be followed, where a frame-wise colorization approach will be employed. This means that the input video will be divided into individual frames, and each of these will be treated as a grayscale image. Unlike the image colorization approach, where a different model was not employed, in this case, a different model will not be employed. Instead, the same model will be reused to colorize each of the frames. This will greatly simplify the system.

Once the color prediction stage is completed, the final stage is the post-processing stage, where the predicted values of the a and b channels are combined with the original L channel to convert the image back to the RGB format. The final

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

colorized image is then displayed to the user, with the option to download the image if required.

## 4. SYSTEM ARCHITECTURE

The system which is proposed is based on an efficient architecture which is designed to perform the process of colorization both for grayscale images as well as videos with the help of deep learning techniques. Moreover, the system is designed to ensure the performance of the process in an efficient manner with the help of the cGAN model

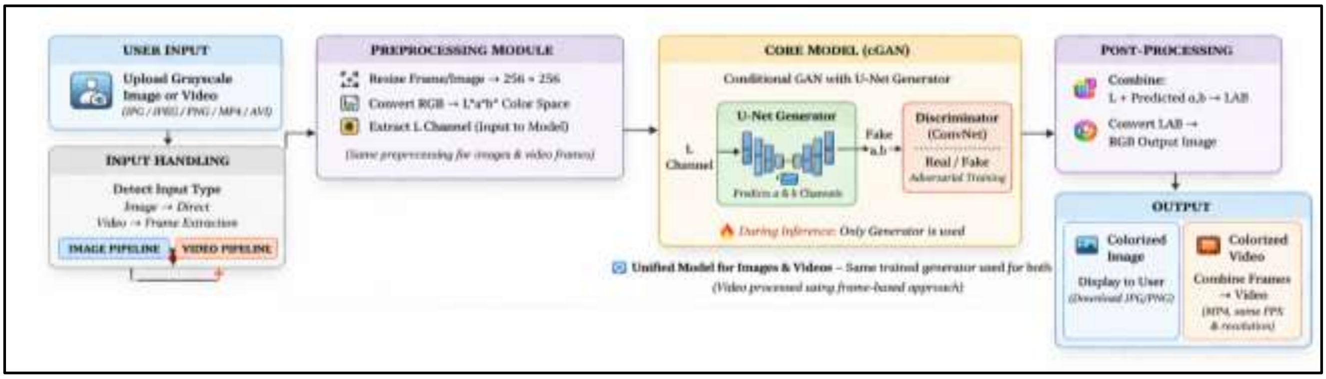


*Fig 4.1: System Architecture*

The architecture which is proposed is based on the creation of different components, including the user interface, the preprocessing module, the colorization model, the video processing module, as well as the output module. Moreover, the user is able to interface with the system with the help of the user interface which is based on the web interface. The user is required to upload the grayscale images as well as the videos.

The input is then fed into the system with the help of the preprocessing module, where the images are resized as well as converted to the Lab color space. The images can be processed with the help of the cGAN model depending on the input which is either an image or a video.

### 4.1 Image Colorization Module

This module is in charge of processing the grayscale images. The preprocessed image is then sent through the cGAN model that has already been trained. The generator in this model will then produce the missing chrominance channels, i.e., a and b, from the luminance channel, i.e., L. The predicted colors are then combined with the original L channel to produce the image in RGB format. This module ensures that the colors produced are contextually correct while maintaining the edges, texture, and other structural features of the original image.

### 4.2 Video Colorization Module

The video colorization module is designed to work with grayscale videos. The method for video colorization, which is presented in this work, is based on the frame-based method. The video is divided into a number of frames, and then each frame is treated as a separate image. The same model, which is used in the image colorization module, is used on these frames. After processing all of these frames, they are stitched together in a sequence to obtain a video. It needs to be noted that since a frame-based method is used, there can be some differences in these frames, especially in a video with moving scenes. However, this makes the process simple.

*Fig 4.2.1: Reconstructed Colorized Video Frames*

### 4.3 User Interface Module

The output module is used for displaying the processed data in a user-friendly way. Once the colorization process is complete, the results can be viewed by using the interface. It is used to view the difference between the input image in grayscale mode and the processed image. It is also possible for the user to download the processed image/video. Lightweight web technology is used to create a user-friendly interface.

## 5. RESULTS

The performance of the system has been tested by considering different images in grayscale format. The images have been used for testing the performance of the system in order to check the efficiency of the system in producing images in real colors. The performance of the system has been satisfactory in producing images in the right colors. At the same time, the structure of the images has also been maintained by the system. In the case of the image colorization task, the performance of the system has been satisfactory in producing images in the right colors. At the same time, the images have been produced in the right colors, especially in the areas where the texture of the images is high.

The performance of the system has been satisfactory in producing images in the right colors for the video colorization task. At the same time, the images have been produced in such a way that smooth video playback can be ensured. However, the system has processed the images individually. Therefore, some changes can be noticed in the images. Still, the performance of the system has been satisfactory.

*Fig 5.1: System Execution and Deployment Output*

As can be seen in the above figure, the execution of the application in the development environment is a success. The output in the terminal is a clear indication that all the required libraries are installed appropriately, and the system is deployed through the web interface. The generated URLs are a clear indication that the application is available for interaction with the user.


*Fig 5.2: Grayscale to Color Image Conversion Result*

The above figure illustrates the output of the image colorization function. From the above figure, it can be noted that the grayscale image is converted into its colored version through the use of the cGAN model. From the above image, it is evident that the system has the ability to add colors to the image.

*Fig 5.3: Frame-Based Video Colorization Result*

This is a figure showing the output of the colorization using the frame-based method. As can be seen in the figure, the frames of the grayscale video were combined in order to produce the final output. It can be noted that the colorization is successful, and this is in line with the proposed method.

## 6. DISCUSSIONS

The experimental results obtained by the proposed system indicate the effectiveness of the unified deep learning approach for colorization of images and videos. The cGAN model is able to learn complex relationships between the gray-scale image and its color information. The major advantage of this system is that it is fully automated and does not require any human intervention. This is a significant advantage over traditional approaches that require human intervention.

The proposed cGAN model is found to be very effective for colorization of images by identifying the context information and colorizing different regions of the image appropriately. The proposed approach is also found to be very effective in colorizing natural features such as sky, vegetation, and water, as these are most commonly represented in an image. The use of the Lab color space is critical in this approach. The use of this color space helps in the separation of luminance and chrominance information and also helps the model focus on colorization without affecting other features such as edges and textures.

From the results obtained in the video colorization process, it can be noted that the frame-based approach can be used to effectively handle video data. The system can process the video as individual frames, and the same model can be used to process each frame. This greatly simplifies the system's architecture. The resulting video is smooth in terms of visual continuity. The changing colors in consecutive frames can be considered smooth.

Some limitations have been noted in the video colorization process. The system can process the video as individual frames. Some color inconsistencies might be noted in the video. This might be more noticeable in dynamic video, especially in cases of high motion. The flickering effect might be noted. Despite the limitations, the quality of the video can be considered satisfactory.

Another interesting thing to note is that, in terms of generalization, the model can function more accurately on images that are similar to those in the data set provided for training, considering aspects such as images. However, in cases where images are not commonly found, there is a possibility that the accuracy of the model in terms of colors might be compromised to a certain level, showing some unrealism. This is where deep learning techniques are heavily dependent on the data provided.

Considering the system level, integrating all these aspects into one umbrella has its own advantages. The user interface provides a smooth user experience, where users can input their data, start processing, and then retrieve the data without the need for the user to be tech-savvy. The use of a single model for handling images and videos also minimizes memory usage.

As far as the performance of the system is concerned, it can be said that the system has shown promise in this regard by maintaining a balance between efficiency in computations and the quality of the output. Moreover, it can be said that due to the support of libraries such as PyTorch and OpenCV, the system has been able to perform efficiently in processing the images. Even though the performance of the system in terms of real-time processing of images is not up to the mark for videos of high resolutions, the performance of the system can be said to be satisfactory for images of moderate sizes.

It can thus be said that the system has been successful in overcoming the challenges related to the colorization of images in shades of grayscale by the use of deep learning techniques. Moreover, it can be said that the system has shown promise in terms of its efficiency in processing images and videos by the use of the unified framework. Even though there are some drawbacks in the current system, it can be said that the system has been successful in achieving its aim by the production of images that are realistic in nature.

## 7. CONCLUSION

This paper has proposed a unified deep learning-based approach to the problem of automatic colorization of both images and videos, where a Conditional Generative Adversarial Network structure, also known as U-Net, has been employed. This system has the potential to colorize images and videos in an efficient manner, considering that the Lab color model has been considered, ensuring that structural details are maintained, along with realistic colorizations.

The novelty of the proposed approach is based on the utilization of a single model to address both image and video colorization. The proposed approach has been considered a frame-based approach to colorization, ensuring that the structure of the proposed system is simple, thus making the efficiency of the proposed approach higher.

This can be seen in the results, which show the efficiency of the system in producing high-quality colorized images and videos with minimal user intervention. Although some limitations, such as frame-wise inconsistency in videos, can be seen, the performance is satisfactory.

In conclusion, the proposed method is effective, automated, and user-friendly for the task of grayscale colorization. It shows the potential of deep learning techniques in the field of media enhancement and opens up the scope for further improvements to achieve greater accuracy and efficiency.

## 8. REFERENCES

[1] Anwar, S., Barnes, N., & Petersson, L., "Image Colorization: A Survey and Benchmark," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025.

[2] Reddy, G. B. S. T., Anudeep, Y., & Kumar, P., "Advanced Adversarial Techniques for Enhanced Grayscale Video Colorization Using Deep Learning," IEEE Conference Proceedings, 2024.

[3] Tadiparthi, M., Arathi, V. N., & Vignesh, V. S., "Gray Scale Image Colorization Using CNN and PyTorch," International Conference on Expert Clouds and Applications, 2024.

[4] Abhiram, K. V., Keerthi, U. S., & Ashok, S., "Parallel Computing and Deep Learning-Based Techniques for Colorization of Visual Media," IEEE International Conference, 2023.

[5] Chatterjee, D., & Iyer, S., "GAN-Based Colorization of Grayscale Images Using Transfer Learning," International Journal of Advanced Computer Vision, 2022.

[6] Singh, P., Kumar, V., & Rao, N., "Automatic Video Colorization with Temporal Consistency Using Deep Neural Networks," Journal of Multimedia Systems, 2023.

[7] Lee, J., Park, H., & Kim, S., "Hybrid CNN-GAN Framework for Enhanced Image Colorization," International Journal of Artificial Intelligence Research, 2024.

[8] Mehta, R., Sharma, P., & Iyer, S., "Advanced Deep Learning Techniques for Realistic Grayscale Image Colorization," Journal of Computer Vision and Image Processing, 2024.

[9] Nguyen, H., Kumar, R., & Lee, H., "Hybrid GAN-Based Framework for Automatic Colorization of Grayscale Images and Videos," International Conference on Artificial Intelligence and Multimedia Systems, 2025.

[10] Zhao, Y., Chen, L., & Wang, M., "Vision Transformer-Based Automatic Image Colorization," IEEE Access, 2023.

[11] Patel, A., & Verma, R., "Deep Learning Approaches for Image and Video Colorization: A Modern Review," Journal of Digital Image Processing, 2022.

[12] Gupta, S., Nair, P., & Reddy, K., "Efficient Color Restoration Using GAN and U-Net Architecture," International Journal of Emerging Technologies, 2024.