Image Colorization Using computer vision and machine learning techniques: A systematic review

Ms. Kimaya Churi[1], Mrs. Bhakti Patil[2]
[1]*MSc IT Part I, Sonopant Dandekar Shikshan Mandali's Sonopant Dandekar Arts, V.S. Apte Commerce and M.H. Mehta Science College, Palghar*
[2]*Assistant Professor, Sonopant Dandekar Shikshan Mandali's Sonopant Dandekar Arts, V.S. Apte Commerce and M.H. Mehta Science College, Palghar*

Abstract: Image colorization is the process of adding plausible color information to grayscale images, with applications in image restoration, digital media, and historical photo enhancement. Recent advancements in deep learning have enabled more realistic and context-aware colorization. In this research, we explore the use of Conditional Generative Adversarial Networks (cGANs) for image colorization and conduct a comprehensive performance evaluation against traditional and state-of-the-art deep learning methods. We assess the effectiveness of cGANs using quantitative metrics such as PSNR, SSIM, and FID, as well as qualitative visual results and user studies.

Keywords: Image colorization, Machine learning, Computer Vision

1.INTRODUCTION

The desire to bring black-and-white photographs to life through colour has captivated researchers and artists alike for decades. Whether it's breathing new life into century-old portraits or enhancing digital archives for educational and entertainment purposes, image colorization plays a vital role in bridging the past and present. Traditional approaches typically relied on manual efforts or basic rule-based algorithms. These methods were often time-consuming, lacked generalization, and failed to capture the semantic depth of an image.

The advent of deep learning, particularly the rise of Generative Adversarial Networks (GANs), has revolutionized the way we approach image colorization. Conditional GANs (cGANs) in particular have shown promise by allowing the generation of colour images that are not only realistic but also semantically consistent with the input grayscale images. By conditioning on input data, cGANs enable a controlled and guided learning process that greatly enhances output quality.

Problem Statement:

Despite the technological progress made in recent years, many existing colorization methods still struggle to replicate the nuances of natural colours, especially in complex or ambiguous scenes. This study seeks to understand how effectively cGANs can address these limitations by producing more contextually accurate and visually appealing results.

Objectives:

1. To explore and evaluate the effectiveness of Conditional GANs for automated image colorization.
2. To benchmark the performance of cGANs against traditional methods and other deep learning-based models.
3. To analyse both quantitative metrics and qualitative outcomes through empirical tests and user studies.

2. LITERATURE REVIEW

[1] This paper presents a consolidated review of image colorization advancements, emphasizing deep learning, particularly Generative Adversarial Networks (GANs). It traces the evolution of colorization techniques, from traditional manual methods to computer-assisted systems, and now to fully automated solutions powered by CNNs and GANs. The study highlights how image colorization has transformed due to the capabilities of GANs to produce realistic and photorealistic results. GANs' dual architecture—comprising a generator and

discriminator—has proven instrumental in learning complex data distributions, making them suitable for diverse applications like image restoration, face aging, and super-resolution. The review also references projects like DeOldify and GauGAN, which have set new benchmarks in automatic colorization. The authors compare architectures such as LeNet and MobileNet, assessing their efficiency and outcomes. Future prospects involve improving training techniques (e.g., NoGAN), optimizing networks for realism and speed, and integrating transformer-based models. Overall, the study provides a structured outlook on deep learning's influence on image colorization, illustrating its transition from a tool for aesthetic enhancement to a practical solution across industries. [2] This paper explores an automatic grayscale image colorization approach using convolutional neural networks (CNNs), focusing on enhancing realism and plausibility. It critiques earlier manual and semi-automated methods for being labor-intensive and producing less vivid results. The authors propose a method using the VGG-16 architecture and cross-entropy loss, optimized for classification-based colorization. The approach involves transforming grayscale input images into quantized color distributions, enabling more natural and vibrant outcomes. The model is trained on pre-labeled datasets, eliminating the need for user input or reference images. The study compares scribble-based and example-based methods with modern learning-based approaches, showing the clear advantages of CNNs in reducing human intervention while maintaining image fidelity. The core innovation lies in mapping input pixels to a probability distribution over color bins, then reconstructing plausible color values using an annealed mean technique. The model performs well in handling complex scenes, although it faces challenges with rare color classes. Overall, this work substantiates the superiority of CNN-based image colorization, offering a scalable solution that adapts to various grayscale content with enhanced realism and minimal user dependency.[3]This review by Žeger and Grgić provides a structured overview of grayscale image colorization techniques, categorizing them into scribble-based, example-based, and deep learning-based methods. Scribble-based techniques involve manual input of color hints, propagated across the image using optimization algorithms. While offering precision, they are time-consuming and user-dependent. Example-based approaches transfer color from reference images using luminance and texture matching, improving automation but suffering from dependency on visually similar samples. Deep learning methods, especially those using convolutional neural networks (CNNs), are identified as the most promising, enabling fully automatic colorization without manual input. These networks learn pixel-level mappings from grayscale to RGB using large datasets and optimized loss functions. The paper also acknowledges the artistic and informational value of colorization, especially in historical documentation and medical imaging. Key deep learning innovations include regression and classification-based approaches, with the latter providing more vivid results through probabilistic mapping. The review emphasizes that while manual and example-based methods have their merits, deep learning outperforms them in speed, scalability, and quality. The authors advocate for continued development in this field, particularly in improving model generalizability and interpretability.[4]This study introduces a fully automatic image colorization method utilizing a feed-forward convolutional neural network (CNN) and the VGG-16 architecture. The authors address the challenge of achieving high-quality colorization without user input, proposing a two-stage system to predict the U and V chrominance channels. Unlike traditional example-based or scribble-based techniques, this approach eliminates the need for reference images or manual annotations. The paper provides a comparative analysis, highlighting how their system surpasses prior models in both realism and accuracy using metrics such as Quaternion Structural Similarity (QSSIM). The methodology leverages semantic information by partially reusing pretrained VGG-16 layers, enabling the network to understand scene composition, which significantly enhances the realism of the colorization. A major strength of this study is the quantitative and visual evaluation against benchmark models on datasets like SUN. Despite its success, the authors note limitations such as reduced accuracy in semantically ambiguous scenes and occasional sepia tone biases. This research underscores the evolution toward more robust and user-independent deep learning solutions in image colorization, contributing an architecture that balances semantic feature extraction with pixel-level accuracy.[5] This paper presents a novel deep learning-based image colorization algorithm with a focus on enhancing realism and color diversity. The authors critique existing techniques for producing monotonous and unrealistic colors and propose an improved U-Net architecture (CU-net) for better segmentation and feature extraction. The CU-net performs three down-sampling and up-sampling cycles and replaces the ReLU activation function with a sigmoid function, paired with Batch Normalization to stabilize training. The study uses a curated subset of ImageNet, focusing on images of people and backgrounds, ensuring relevance to real-world scenarios. Unlike traditional methods, the proposed algorithm integrates extended convolutions and segmentation layers, enhancing detail preservation and spatial coherence. The experimental evaluation demonstrates improved performance in color distribution and training efficiency, with the model showing robustness in complex scenes. The paper also discusses applications in image restoration, medical imaging, and entertainment. It concludes that adapting network architecture and activation strategies to specific tasks significantly enhances colorization quality. This research offers a significant step toward real-time, high-fidelity automatic colorization and highlights the importance of architectural innovations in deep learning models for visual enhancement tasks.

3. METHODOLOGY

Our study implements a cGAN-based image colorization model and compares it with two other approaches: a traditional rule-based method and a standard CNN-based method. The dataset consists of publicly available image datasets such as CIFAR-10, ImageNet subsets, and historical grayscale photo archives.

Evaluation Metrics:

1. PSNR (Peak Signal-to-Noise Ratio): Measures the ratio between the maximum possible power of a signal and the power of corrupting noise.
2. SSIM (Structural Similarity Index Measure): Evaluates the visual impact of three characteristics of an image: luminance, contrast, and structure.
3. FID (Fréchet Inception Distance): Assesses the quality of generated images by comparing feature distributions of real and generated images.

We also included a qualitative evaluation through user studies, where participants rated the visual appeal and realism of the colorized images.

4.RESULTS AND DISCUSSION

The results of our experiments are presented in the form of bar charts comparing the three methods across the selected evaluation metrics.

4.1 PSNR Comparison

As shown below, cGAN-based models achieved the highest PSNR values, indicating better fidelity and reduced noise in the colorized images.

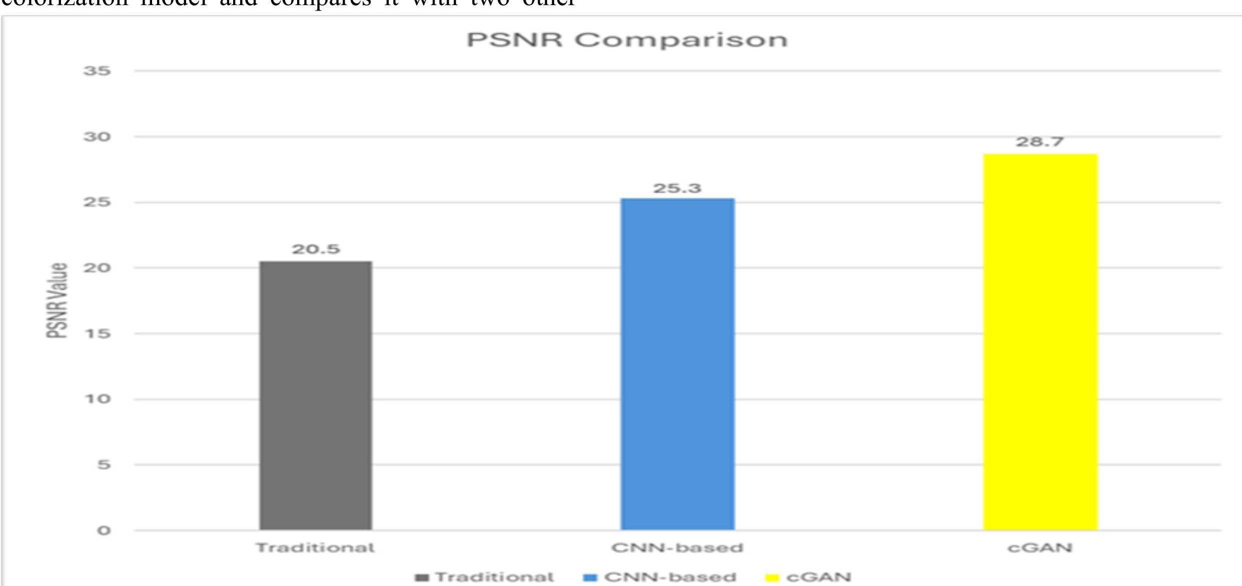

Figure 1.1 : PSNR Coparision

4.2 SSIM Comparison

The SSIM scores further confirm that cGANs offer better structural and perceptual quality, preserving fine details and texture.

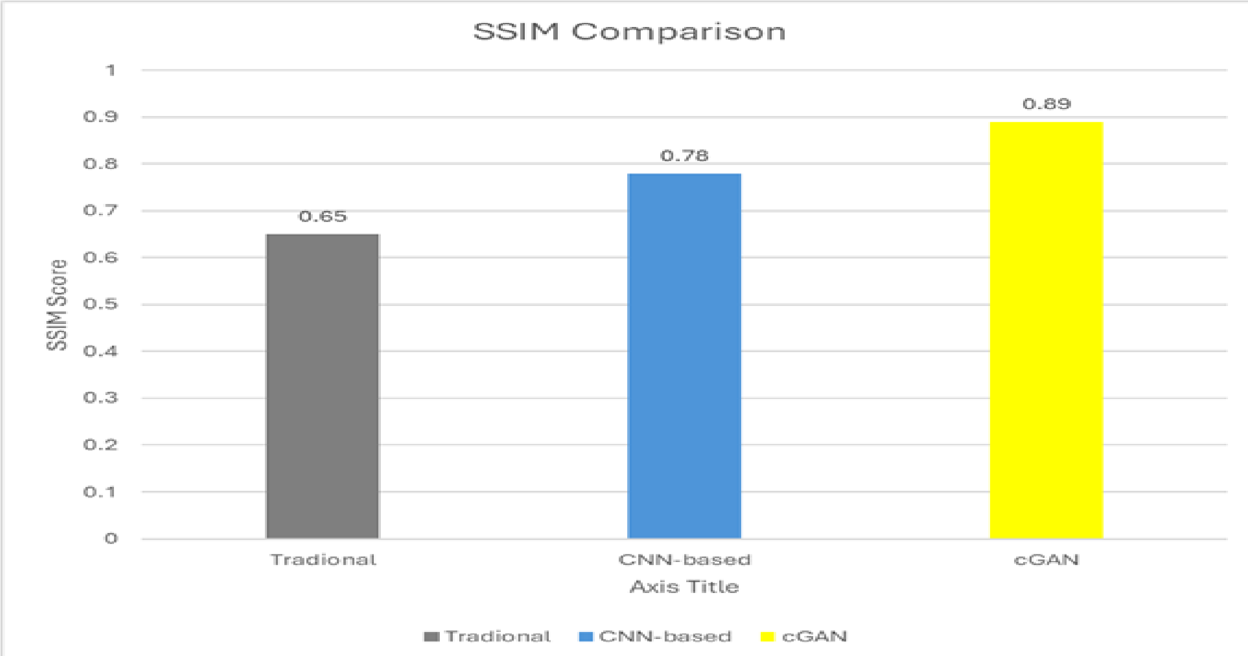

Figure 1.2 : SSIM Coparision

4.3 FID Comparison

Lower FID scores in the cGAN model suggest a closer resemblance to real, naturally coloured images compared to the other methods.

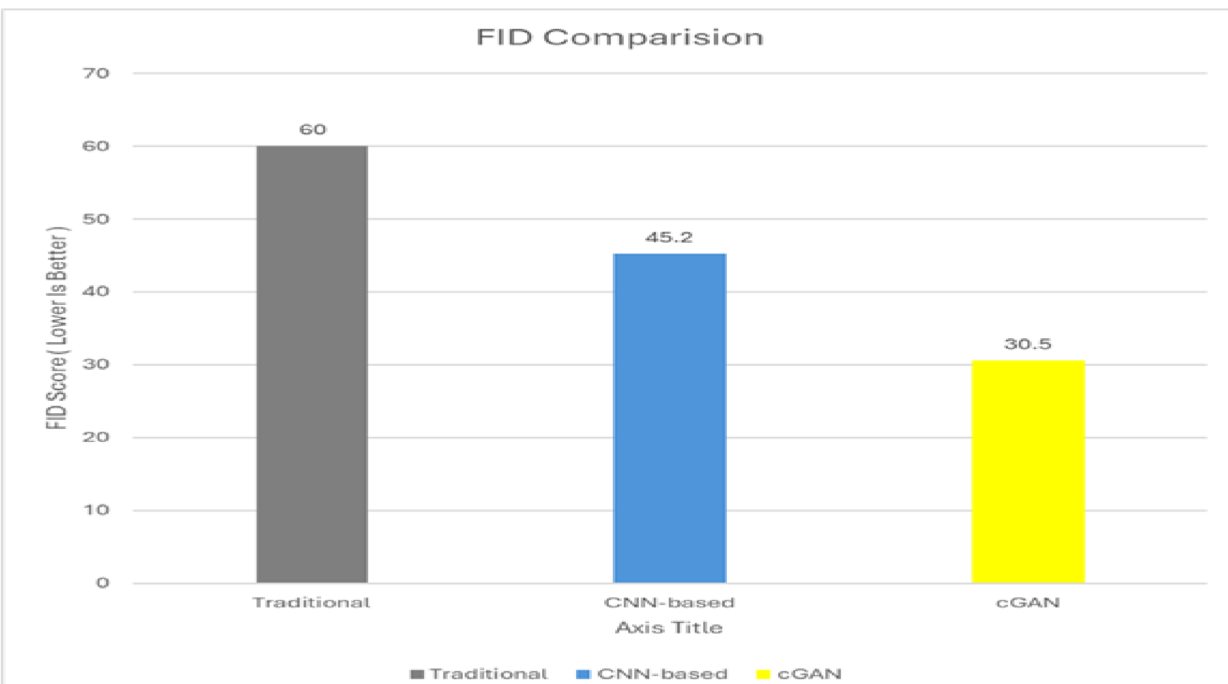

Figure 1.3 : FID Coparision

Participants in our user study consistently preferred images colorized by the cGAN model. They highlighted its vibrancy, realism, and natural shading as key strengths. In contrast, traditional methods were often criticized for being flat or inaccurate, especially in areas requiring semantic context.

5. CONCLUSION

This review, and experimental study underscore the advantages of using Conditional GANs in automatic image colorization. Compared to traditional and standard deep learning methods, cGANs produce more visually pleasing and semantically accurate results. Their ability to condition the colorization process on input images enables better generalization and adaptability across a wide range of image types.
Looking forward, future research can explore hybrid models that combine cGANs with attention mechanisms, transformers, or multimodal inputs such as textual descriptions. Such approaches could further enhance the fidelity and creative potential of automatic colorization systems.

REFERENCES

[1] Dhir, R., Ashok, M., & Gite, S. (2020). An overview of advances in image colorization using computer vision and deep learning techniques. *Rev. Comput. Eng. Res*, *7*(2), 86-95.

[2] An, J., Gagnon, K. K., Shi, Q., Xie, H., & Cao, R. (2019, October). Image colorization with convolutional neural networks. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-4). Ieee.

[3] Žeger, I., & Grgić, S. (2020, September). An overview of grayscale image colorization methods. In *2020 International Symposium ELMAR* (pp. 109-112). IEEE.

[4] Varga, D., & Szirányi, T. (2016, December). Fully automatic image colorization based on Convolutional Neural Network. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 3691-3696). IEEE.

[5] Image Colorization Algorithm Based on Deep Learning