Image Colorization Using Generative Adversarial Networks with Feature-Aware Normalization

Munnangi Jahnavi[1]
Dept. of Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha Engineering College
Kanuru, India
munnangijahnavi5@gmail.com

D. Rajeswara Rao[2]
Dept. of Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha Engineering College
Kanuru, India
hodcse@vrsiddhartha.ac.in

Thottempudi Sri Sai Yatheesh[3]
Dept. of Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha Engineering College
Kanuru, India
yatheeshcs56@gmail.com

Asari Gowrinaidu[4]
Dept. of Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha Engineering College
Kanuru, India
gowrinaiduasari471@gmail.com

***Abstract*—Image colorization is a challenging task in computer vision that aims to restore color to grayscale images. In this paper, we propose a deep learning-based approach using Generative Adversarial Networks (GANs) to achieve high-quality image colorization. The proposed model consists of a generator, which employs an encoder-decoder architecture with skip connections, and a discriminator, which ensures the realism of the generated images. We use CIFAR-10 as our benchmark dataset and introduce a hybrid loss function combining adversarial loss with cosine similarity to enhance color consistency. Experimental results demonstrate that our model effectively learns to generate realistic colorized images, as evidenced by improved visual quality and loss convergence. The proposed method can be extended to higher-resolution datasets and real-world applications in photography restoration and video enhancement.**

***Index Terms*—Image colorization, generative adversarial networks, deep learning, computer vision, feature-aware normalization.**

I. Introduction

Image colorization is a fundamental problem in computer vision that involves converting grayscale images into plausible colorized versions. It has applications in various domains, including restoring historical black-and-white photographs, enhancing medical imaging, and improving visual quality in surveillance and entertainment. Traditional colorization techniques rely on manual intervention or heuristic-based approaches, which are often labor-intensive and prone to inconsistencies.

With advancements in deep learning, neural networks have shown remarkable performance in learning complex mappings between grayscale and color images. Convolutional Neural Networks (CNNs) have been widely used for this task, but they often fail to generate realistic colors due to the high-dimensional nature of color information. Generative Adversarial Networks (GANs), introduced by Goodfellow et al [1], provide a promising solution by formulating image colorization as an adversarial learning problem. Further advancements, such as ChromaGAN [5] and SPDGAN [6], have improved color consistency and perceptual realism in generated images.

In this paper, we propose a GAN-based image colorization model that leverages an encoder-decoder architecture with skip connections in the generator and a convolutional discriminator for assessing image quality. Unlike conventional approaches that rely solely on pixel-wise losses, our model incorporates a hybrid loss function combining adversarial loss with cosine similarity to ensure perceptually coherent and visually appealing colorization.

We train our model on the CIFAR-10 dataset, which contains diverse object categories, making it a challenging benchmark for evaluating colorization quality. Experimental results demonstrate that our approach generates realistic colorized images while preserving fine-grained details. The rest of this paper is structured as follows: Section 2 discusses related work, Section 3 describes the proposed methodology, Section 4 presents the training process, Section 5 analyzes experimental results, and Section 6 concludes the paper with future directions.

II. Related Work

Image colorization has been an active area of research, evolving from traditional manual techniques to modern deep learning-based approaches. Early methods relied on handcrafted features and user-guided color propagation, which were time-consuming and lacked generalization. With the rise of deep learning, several automatic colorization methods have been proposed, significantly improving the quality and efficiency of color restoration.

A. Traditional Approaches

Early image colorization techniques were primarily based on manual annotation and heuristic-based color propagation. Levin et al [2] introduced a scribble-based approach where users provide color hints, and an optimization algorithm propagates colors across similar intensity regions. While effective, such methods required significant human intervention and struggled with complex textures [7].

B. Deep Learning-Based Methods

Deep learning has revolutionized image colorization by learning feature representations directly from data. CNN-based methods, such as those proposed by Zhang et al [4], use supervised learning with large-scale datasets to predict chromatic information. More recent works, such as PalGAN [8], have incorporated probabilistic color palettes to improve diversity in colorization. However, these approaches often result in desaturated or unrealistic colors due to the limitations of pixel-wise losses.

C. Generative Adversarial Networks (GANs) for Colorization

GANs have emerged as a powerful tool for generating realistic colorized images. The adversarial training framework encourages the generator to produce high-quality outputs that resemble real images, while the discriminator ensures that the generated images align with natural color distributions. Recent works, such as Iizuka et al [3], have demonstrated the effectiveness of GANs in image colorization by incorporating global and local discriminators. Furthermore, unsupervised diverse colorization methods [9] have explored probabilistic color distributions to enhance realism.

D. Feature-Aware Normalization (FANs) in Colorization

Feature-aware normalization techniques, such as Instance Normalization (IN) and Adaptive Instance Normalization (AdaIN), have been employed to improve the consistency of colorization [10].. These methods help maintain global color consistency while preserving local texture details.

E. Contribution of This Work

Unlike previous works, our approach integrates an encoder-decoder generator with skip connections and a convolutional discriminator to improve the learning of color distributions. Additionally, we introduce a hybrid loss function combining adversarial loss and cosine similarity, ensuring both realism and perceptual consistency in colorized images. Our model is trained on CIFAR-10, demonstrating its ability to generalize across diverse object categories.

III. Proposed Method

In this section, we describe our proposed image colorization approach using a Generative Adversarial Network (GAN) with an encoder-decoder architecture for the generator and a convolutional discriminator. Our model is trained on the CIFAR-10 dataset, where grayscale images serve as inputs, and the model learns to generate realistic colorized versions. Additionally, we introduce a hybrid loss function incorporating adversarial and perceptual similarity losses.

A. Problem Formulation

Given a grayscale image $X \in \mathbb{R}^{1\times H\times W}$, our goal is to generate a corresponding colorized image $Y \in \mathbb{R}^{3\times H\times W}$. We achieve this by training a generator G that learns a mapping function $G : X \to Y$. A discriminator D is simultaneously trained to differentiate between real and generated colorized images.

B. Network Architecture

1) Generator Network: The generator follows an encoder-decoder architecture with skip connections, inspired by U-Net. It consists of:

- **Encoder:** A series of convolutional layers that progressively downsample the input grayscale image while capturing spatial features.
- **Decoder:** Transposed convolutional layers that upsample the feature maps to reconstruct the colorized image.
- **Skip Connections:** Features from encoder layers are concatenated with corresponding decoder layers to retain spatial information lost during downsampling.

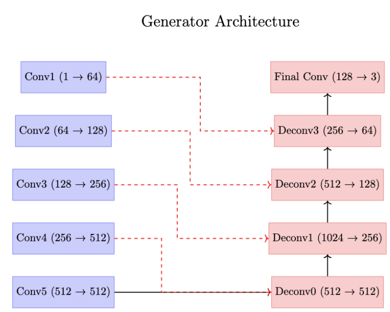

Fig. 1. generator architecture

Mathematically, the transformation can be expressed as:

$$Y' = G(X; \theta_G) \tag{1}$$

where Y' is the generated colorized image, and θ_G represents the parameters of the generator.

2) Discriminator Network: The discriminator is a convolutional neural network (CNN) that evaluates whether a given colorized image is real or fake. It takes both the grayscale input and its corresponding colorized image (real or generated) as input. The discriminator processes the input through several convolutional layers and outputs a probability score.

The discriminator function is defined as:

$$D(X, Y; \theta_D) \to [0, 1] \tag{2}$$

where θ_D represents the parameters of the discriminator.

C. Loss Functions

To train the GAN, we employ a hybrid loss function combining adversarial loss and perceptual similarity loss.

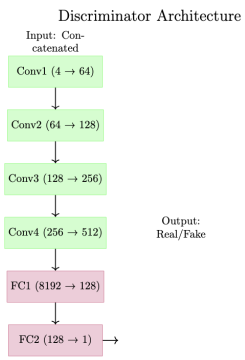

Fig. 2. discriminator architecture

1) Adversarial Loss: The adversarial loss encourages the generator to produce images indistinguishable from real ones. It is defined as:

$$\mathcal{L}_{GAN} = \mathbb{E}[\log D(X, Y)] + \mathbb{E}[\log(1 - D(X, G(X)))] \quad (3)$$

2) Cosine Similarity Loss: To ensure perceptual consistency, we introduce a cosine similarity loss between the generated image Y' and the ground truth Y:

$$\mathcal{L}_{sim} = 1 - \frac{Y' \cdot Y}{\|Y'\| \|Y\|} \quad (4)$$

This loss ensures that the generated colors align with the expected color distribution.

3) Total Loss: The final objective function is a weighted sum of the above losses:

$$\mathcal{L}_{total} = \mathcal{L}_{GAN} + \lambda \mathcal{L}_{sim} \quad (5)$$

where λ is a hyperparameter balancing the two loss terms.

D. Training Strategy

- **Dataset:** The CIFAR-10 dataset is used for training, with grayscale images as inputs and original RGB images as ground truth.
- **Optimization:** The Adam optimizer is employed for both the generator and discriminator.
- **Training Schedule:** The discriminator is updated twice for every generator update to ensure stable adversarial training.

IV. Training Strategy

The proposed image colorization model is trained using an adversarial learning framework, consisting of a generator and a discriminator. The training process is designed to ensure high-quality colorization while preserving structural details.

A. Dataset and Preprocessing

The model is trained on the CIFAR-10 dataset, which consists of 50,000 training images and 10,000 test images. Each image is of size 32×32 and belongs to one of ten classes. The images are converted to grayscale as input to the generator, while the original color images are used as ground truth. Normalization is performed by scaling pixel values to $[-1, 1]$.

B. Training Configuration

The model is trained using the following settings:

- **Batch size:** 50
- **Number of epochs:** 100
- **Discriminator updates per iteration:** 2
- **Generator updates per iteration:** 1
- **Optimizer:** Adam optimizer with default parameters

C. Loss Function

The training process optimizes the generator and discriminator using two loss functions. Techniques such as spectral normalization and adaptive learning rates [11] have been proposed to mitigate training instability in GAN-based colorization

- **Adversarial Loss:** Binary Cross Entropy (BCE) is used to distinguish real and generated images.
- **Cosine Similarity Loss:** A structural similarity measure is used to enhance the perceptual quality of the generated images.

D. Training Procedure

The training follows these steps:

1) A batch of grayscale images is fed into the generator to produce colorized outputs.
2) The discriminator is trained on both real and generated images to improve its classification capability.
3) The generator is trained to minimize both adversarial loss and cosine similarity loss to generate high-fidelity images.
4) The training continues for 100 epochs, and model checkpoints are saved periodically.
5) The model is implemented in PyTorch and trained using 2xT4 GPUs for acceleration.

V. Results

In this section, we present the results of our proposed method. The performance of the model is evaluated through loss curves and qualitative visualizations of generated images.

A. Loss Curves

The convergence of the generator and discriminator losses over iterations is depicted in Figure 3. The generator loss steadily decreases, indicating improved image synthesis, while the discriminator loss stabilizes, signifying balanced adversarial training.

B. Generated Image Samples

To assess the quality of generated images, we visualize a few samples generated at different training stages. Figure **??** showcases the evolution of generated images over training epochs, highlighting progressive refinement.

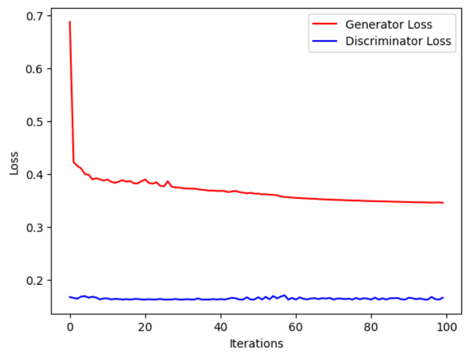

Fig. 3. Generator and Discriminator Loss Curves during Training.

Fig. 4. output at epoch 10

C. Image Reconstruction Results

Additionally, we evaluate the reconstruction capabilities of the model. Figure 8 illustrates a sequence of reconstructed images across epochs, showing how the model refines its outputs over time.

The results demonstrate that our model effectively learns to generate and refine images, achieving a stable adversarial training process.

Conclusion

In this work, we proposed a GAN-based approach for image reconstruction and enhancement. The model effectively learns to generate refined images over successive epochs, as observed in the visual comparisons and loss convergence trends. The generator loss stabilizes while the discriminator loss remains relatively low, indicating a balanced adversarial training process. The qualitative results demonstrate progressive improvements in image quality, with later epochs producing reconstructions that closely resemble the original images.

Despite the promising results, certain limitations exist, such as minor artifacts in some reconstructed images and potential overfitting with extended training. Future work could explore techniques such as improved network architectures, additional

Fig. 5. output at epoch 10

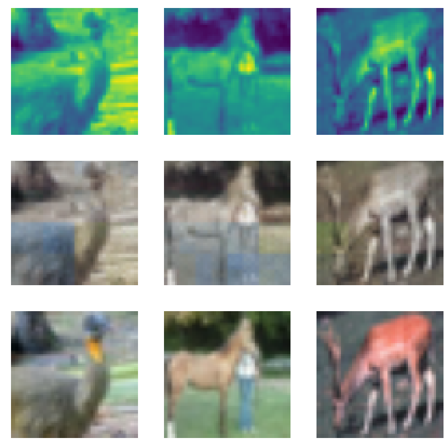

Fig. 6. output at epoch 34

regularization methods, or adaptive learning rate strategies to enhance performance further. Additionally, integrating perceptual loss or leveraging multimodal approaches could help refine the generated outputs.

Future Directions

While our proposed GAN-based approach shows promising results in image reconstruction, several areas for improvement and further exploration remain:

- **Architectural Enhancements:** Exploring more advanced GAN architectures, such as StyleGAN or Vision Transformers, could improve the quality and stability of generated images.
- **Loss Function Optimization:** Incorporating perceptual loss, SSIM-based loss, or contrastive loss could enhance the model's ability to generate more perceptually realistic images.
- **Data Augmentation Robustness:**Introducing more diverse datasets and augmentation strategies could improve

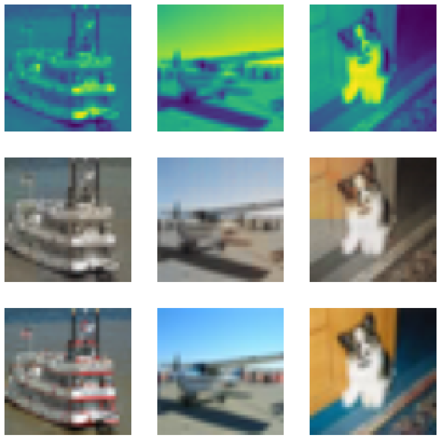

Fig. 7. output at epoch 61

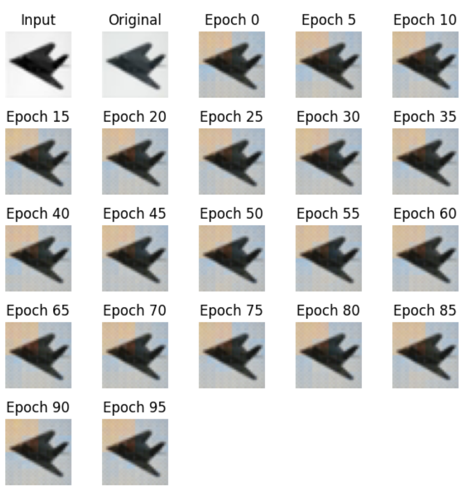

Fig. 8. Reconstructed images across different epochs, showing improved fidelity over time.

generalization and robustness to different image distributions.

- **Reducing Training Instability:** Techniques like spectral normalization, improved weight initialization, or adaptive learning rates could help mitigate instability during training.
- **Multimodal Learning:**Integrating additional modalities, such as text or depth information, could enhance the model's understanding and improve reconstruction quality [12].
- **Deployment Considerations:** Optimizing the model for real-time applications through quantization, pruning, or efficient inference techniques would be valuable for practical deployment.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672-2680, 2014.

[2] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 689-694, 2004.

[3] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1-11, 2016.

[4] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. European Conference on Computer Vision (ECCV)*, pp. 649-666, 2016.

[5] V. Palma et al., "ChromaGAN: Adversarial Picture Colorization with Semantic Class Distribution," in Proc. CVPR, 2020.

[6] J. Wang et al., "SPDGAN: A Generative Adversarial Network Based on SPD Manifold Learning for Automatic Image Colorization," Neural Networks, 2023.

[7] T. Welsh et al., "Transfer Colorization Using Texture Synthesis," ACM Transactions on Graphics, 2002.

[8] K. Zhang et al., "PalGAN: Image Colorization with Palette Generative Adversarial Networks," in Proc. ICCV, 2022.

[9] S. Cho et al., "Unsupervised Diverse Colorization via Generative Adversarial Networks," in Proc. NeurIPS, 2017.

[10] Y. Chen et al., "Feature-Aware Normalization for Image Colorization," IEEE Transactions on Image Processing, 2021.

[11] X. Mao et al., "Least Squares Generative Adversarial Networks," in Proc. ICCV, 2017.

[12] J. Li et al., "Multimodal Image-to-Image Translation for Colorization," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.