ColorSphere: A Transfer Learning Approach to Image Colorization Excellence

Dr. David Raj Micheal
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
davidraj.micheal@vit.ac.in

Akhila Raveendran P M
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
akhila.raveendran2023@vitstudent.ac.in

***Abstract*—Image colorization is the process of assigning colors to a grayscale image, transforming it into a visually plausible and perceptually meaningful color image. This technique is vital for restoring historical photos, enhancing medical images, and improving the visual quality of old, damaged images. This paper develops an image colorization approach using the ECCV16 and SIGGRAPH17 models, both of which leverage convolutional neural networks and transfer learning to achieve high-quality colorizations. By utilizing the LAB colorspace instead of RGB, the models achieve more realistic and vibrant colorization. This approach highlights the effectiveness of transfer learning and deep learning techniques in enhancing image colorization.**

***Index Terms*—Image colorization, Neural Networks, Deep Learning, Encoder, Decoder.**

I. INTRODUCTION

Image colorization, the process of adding color to grayscale images, has become increasingly important due to its wide range of applications and advancements in deep learning techniques.Beyond aesthetics, image colorization plays a crucial role in restoring and enhancing historical photographs and videos, allowing us to connect with the past in a vivid and engaging way. In the medical field, colorizing monochrome images from MRI, CT scans, and X-rays can improve interpretability and diagnostic accuracy. Additionally, automatic colorization can enhance the visual quality of old, damaged images, making them more appealing and informative.This paper explains how transfer learning and deep learning approaches significantly accelerated the progress in image colorization. By leveraging pre-trained models , which capture rich feature representations from large datasets, these techniques reduce training time and improve accuracy. This enables the development of sophisticated colorization models that produce more realistic and efficient results, making the process faster and more accessible for various applications.

II. OBJECTIVE

The primary objective of this project is to develop a model capable of automatically colorizing grayscale images by predicting the color information for each pixel. Leveraging the pre-trained models through transfer learning, the project aims to enhance the performance and efficiency of the colorization process.The importance of image colorization lies in its wide range of applications and benefits. Colorized images can significantly enhance visual appeal and provide more information, making them useful in various fields such as historical photo restoration, medical imaging, and film industry
This paper demonstrate the application of deep learning techniques, particularly convolutional neural networks (CNNs), in solving the image colorization problem.Transfer learning has revolutionized image colorization by utilizing pre-trained models , which capture rich features from large datasets. This approach reduces training time and enhances accuracy, enabling more realistic and efficient colorization of grayscale images.

III. LITERATURE REVIEW

Image colorization is a challenging problem in computer vision that involves assigning colors to grayscale images to enhance their visual appeal and perceptual quality. Recent advancements in deep learning, particularly convolutional neural networks (CNNs) and generative adversarial networks (GANs), have significantly improved the accuracy and realism of automated colorization techniques.
Liu et al. [1] in their paper presents a novel automatic grayscale image colorization method based on histogram regression. It uses a source image for color information and performs locally weighted regression on both the grayscale and source images to obtain feature distributions. A new matching method aligns these features by adjusting the histogram's zero-points, achieving luminance-color correspondence in a weighted manner. Additionally, a new evaluation method assesses the confidence of the colorization results. Various experiments demonstrate the method's validity.The papaer by Cheng et al. [2] proposes a fully-automatic method for converting grayscale images to color using deep learning and a large-scale reference database. Unlike traditional methods requiring manual adjustments, this approach ensures high-quality, artifact-free results with a joint bilateral filtering post-processing step. Experiments show it outperforms current algorithms in quality and speed.Suarez et al. [3] propose a deep learning method for colorizing infrared images using a convolutional neural network (CNN). The network, trained on paired infrared and color images,

predicts color from infrared input. By fine-tuning a pre-trained network, the method captures complex relationships between infrared intensities and visible colors, resulting in accurate and natural-looking colorized images..Previous colorization methods often produce a single, most likely solution.But Deshpande, Aditya, et al. [4] generates multiple plausible colorizations using a variational autoencoder and build a conditional model for the multi-modal distribution between grey-level image and the color field embeddings. This approach ensures diversity and quality, outperforming existing techniques.The paper by Varga et l. [5] presents a method for colorizing grayscale images using two parallel Convolutional Neural Networks (CNNs) and a reference image. One CNN uses the reference image to aid color prediction, while the other identifies key areas in the input image. Experiments on the SUN database and other images were evaluated using Peak Signal-to-Noise Ratio (PSNR) and Quaternion Structural Similarity (QSSIM). Nazeri et al., [6] in their paper aims to generalize the image colorization process using a conditional Deep Convolutional GAN (DCGAN).The network is trained on publicly available datasets like CIFAR-10 and Places365.Previous manual methods for colorizing grayscale images often resulted in desaturated colors.The paper by Jiaojiao Zhao et.al [7] explores a novel approach to image colorization by incorporating pixel-level object semantics1. The authors propose a hierarchical neural network with two branches: one for learning object semantics and the other for learning object colors. This method aims to address common issues in image colorization, such as context confusion and edge color bleeding, by optimizing both semantic segmentation and colorization losses.Inspired by recent advancements An, Jiancheng, et al. [8] developed an automatic approach using deep neural networks. By experimenting with various models and loss functions,it is found that training a VGG-16 CNN with cross-entropy loss produced realistic colorizations.Ozbulak et.al [9] in their paper explores the use of Capsule Networks (CapsNet) for the task of image colorization. By using deeper layers from a pre-trained VGG-19 model and modifying the loss function to Mean Squared Error (MSE), the Colorizer Capsule Network (ColorCapsNet) achieves promising results on the DIV2K dataset. The paper by Su et.al [10] presents a novel method for colorizing images with multiple objects. It employs an object detector to isolate object images and an instance colorization network to extract object-level features. These features are then combined with image-level features through a fusion module to predict the final colors. This approach significantly improves colorization quality, outperforming existing methods on various metrics and achieving state-of-the-art performance.The paper by Wu, Yanze, et al. [11]leverages pretrained Generative Adversarial Networks (GANs) to enhance image colorization. By retrieving matched features via a GAN encoder and incorporating them into the colorization process, the method produces vivid and diverse colors. It also allows for controllable and smooth transitions by modifying GAN latent codes, achieving superior performance compared to previous methods.The method introduced by Wu, Min, et al. [12] for colorizing remote sensing images uses a deep convolutional GAN with a symmetrical auto-encoder structure and a multi-scale convolutional module to retain image features during up-sampling and down-sampling. The discriminator, based on ResNet-18, competes with the generator for effective optimization. The method converts images from RGB to YUV, uses the Y channel as input to predict UV channels, and then combines them back into RGB. Experiments show that this method performs well in both visual quality and objective metrics.To tackle multimodal ambiguity and color bleeding in image colorization, the paper by Wang, Yi, et al. [**?**] proposes PalGAN, a GAN-based approach. It estimates a probabilistic palette from the grayscale image and assigns colors using a generative model. Chromatic attention handles color bleeding by considering semantic and intensity correlations. PalGAN outperforms existing methods, delivering diverse, contrastive, and edge-preserving results, and enables color transfer between images with different contexts.Wu, Di, et al. [14] proposes method enhances grayscale image colorization of ethnic costumes by using fine-grained semantic information with Pix2PixHD.The network uses the combined fine-grained semantic and grayscale images as input for generative adversarial networks (GANs).The SCSNet method proposed by Zhang, Jiangning, et al. [15]efficiently combines image colorization and super-resolution into a single process. It features a colorization branch with a Pyramid Valve Cross Attention (PVCAttn) module for aggregating feature maps and a super-resolution branch with a Continuous Pixel Mapping (CPM) module for predicting high-resolution images. SCSNet supports both automatic and referential modes, demonstrating superior performance over state-of-the-art methods with lower FID scores, fewer parameters, and faster running speeds.The paper by Wang et al. [16] introduces a CNN-based method for realistic and diverse image colorization of human images. Using a dataset of 5000 images, the authors enhance the U-net network with improved down-sampling, up-sampling, expanded convolutions, and sigmoid activation with Batch Normalization. This approach reduces training time and improves segmentation quality.Treneska, Sandra, et al. [17]proposes an image colorization method using conditional GANs (cGANs) for image colorization as a self-supervised learning task to avoid manual annotation. The method leverages the realistic colorization capabilities of GANs and transfers the learned features to improve performance in multilabel image classification and semantic segmentation. Experiments on the COCO and Pascal datasets show a 5 percent improvement in classification and 2.5 percent in segmentation, demonstrating the effectiveness of this approach.Kang, Xiaoyang, et al. [18]introduces a method using dual decoders to enhance image colorization. It employs a pixel decoder for spatial resolution and a query-based color decoder for refining colors, reducing color bleeding and enhancing color richness.SS-CycleGAN is a novel image colorization method introduced by Li, Bin, et al.

[19] that ensures semantic and spatial rationality. It uses a self-attention patch discriminator and a Multi-scale Cascaded Dilated Convolution (MCDC) module to maintain detail and spatial consistency. The method outperforms state-of-the-art techniques on the Natura Color and Flower datasets, producing higher quality colorized images.The method proposed by Ambadkar et al. [20] for grayscale image colorization is a 38-layer deep convolutional residual network using the CIELAB color space to reduce solution complexity. It includes 16 residual blocks with 128 convolutional filters and 4 convolutional blocks for image reconstruction. Experiments on Imagenet, Intel, and MirFlickr datasets show significant generalization and high performance in PSNR, SSIM, and PIQE metrics. The method is simpler (16 million parameters), faster (15 images/sec), and more resource-efficient (50,000 training images) compared to state-of-the-art techniques.

IV. METHODOLOGY

A. Preprocessing

This involves preparing the input grayscale image for the deep learning models. These steps are essential to ensure that the image is in the correct format and scale, and that it matches the following input requirements of the pre-trained ECCV and SIGGRAPH models.

1.Input Image Size:

a) Fixed Size: Both ECCV and SIGGRAPH models expect the input image to be resized to 256x256 pixels. This size standardizes the input for these models, which do not have resizing layers within their architectures.

b) Single Channel (Grayscale Lightness Channel): Only the lightness (L) channel is input to the model. The A and B color channels are not provided initially, as the model is designed to predict these color channels.

2. Color Space

a) LAB Color Space: The models are trained in the LAB color space. The LAB color space separates lightness (L) from color information (A and B channels). The input to the model is only the L channel (grayscale lightness), while the model's goal is to predict the missing A and B color channels.

b) Normalized L Channel Range: The L channel should be normalized to the range of -50 to 50. This normalization centers the pixel values around zero, which helps the model in stabilizing predictions and matches the training setup for both ECCV and SIGGRAPH models.

B. Model Selection

Two different models are utilized for predicting the AB channels: one based on the ECCV (European Conference on Computer Vision) approach and the other based on SIGGRAPH (Special Interest Group on Computer GRAPHics and Interactive Techniques). Both models are pre-trained on large datasets of color images, allowing them to learn the relationships between grayscale images and their color counterparts.

ECCV Model Architecture

1) U-Net Structure: The ECCV model is based on the U-Net architecture, a neural network specifically designed for tasks that require precise predictions at the pixel level, such as colorizing grayscale images.

2) Two Parts: The U-Net consists of two main components: an encoder and a decoder. The encoder reduces the image size while extracting key features, allowing the model to focus on important details. The decoder then expands the image back to its original dimensions while incorporating the extracted features to add color.

3) Skip Connections: A key feature of the U-Net is the skip connections that link corresponding layers of the encoder and decoder. These connections allow the decoder to access information from the encoder, ensuring that fine details are preserved during the colorization process. This approach enhances the model's ability to produce sharp and accurate colorized images.

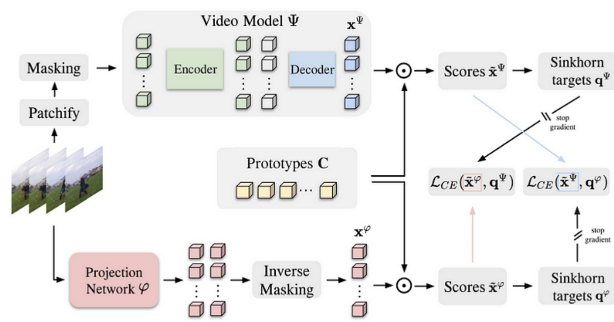

Fig. 1. architecture of ECCV

SIGGRAPH Model Architecture

4) Generative Adversarial Networks (GANs): The SIGGRAPH model commonly utilizes Generative Adversarial Networks (GANs), which consist of two neural networks that compete with each other. One network is called the **generator**, and its role is to create a colored image from a black-and-white input. The other network is the **discriminator**, which evaluates whether the colorized image produced by the generator appears realistic.

5) Generator: The generator processes the grayscale image and applies techniques similar to those used in the U-Net structure or other neural network architectures to add color to the image.

6) Discriminator: The discriminator's function is to determine if the generator's output looks real or fake. This adversarial relationship between the two networks drives the generator to enhance its coloring abilities over time, aiming to produce more convincing results.

7) Loss Functions: The SIGGRAPH model employs specific metrics known as **loss functions** to assess the performance of the generator:

- **Adversarial Loss:** This metric encourages the generator to create images that are realistic enough to fool the discriminator.
- **Content Loss:** This measure ensures that the generated colors align with what should be present by comparing the colorized output to actual color images.
- **Color Loss:** This directly evaluates how closely the predicted colors match the actual colors in the images.

8) Multi-Scale Processing: Some versions of the SIGGRAPH model analyze images at various scales. By looking at the image in different sizes, the model can capture both large and small details more effectively, leading to improved final colorized outputs.

9) Output: Similar to the ECCV model, the SIGGRAPH model generates the A and B channels representing color information, which can then be used to convert the grayscale image into a full-color image.

10) Convolutional Layers: The model employs convolutional layers, which are designed to recognize patterns within the image. These layers analyze small regions of the image one at a time, effectively capturing spatial information.

11) Output: At the end of the process, the model generates the color information (the A and B channels) for the grayscale image. This predicted color information is subsequently transformed back into RGB format to yield a colorized version of the original grayscale image.

C. Forward Pass through the Models

The normalized L channel tensor is passed through the ECCV and SIGGRAPH models individually. This process is referred to as a forward pass, where the model processes the input to produce predictions:

ECCV Model: This model employs a specific architecture (such as a convolutional neural network) optimized for colorization tasks. It predicts the AB channels by learning from the patterns and structures in the training data.

SIGGRAPH Model: Similarly, this model utilizes a different architecture or technique that may yield complementary or improved results based on its design choices.

During this step, the models leverage learned weights and biases to compute the output based on the input L channel. The output is a tensor representing the predicted AB channels, typically in the range of [-128, 127], as the LAB color space has a different range for the A and B channels.

V. OBSERVATIONS

When comparing the colorized images, the output from the SIGGRAPH 17 method (Fig. 2) appears to be more accurate and visually appealing. The apple in the SIGGRAPH 17 output closely resembles the original image with its red color and green leaf, whereas the ECCV 16 output has a mix of green and brown hues that don't match the original colors as well. The SIGGRAPH 17 method seems to better capture the true colors and details, making it the more effective colorization technique in this case.

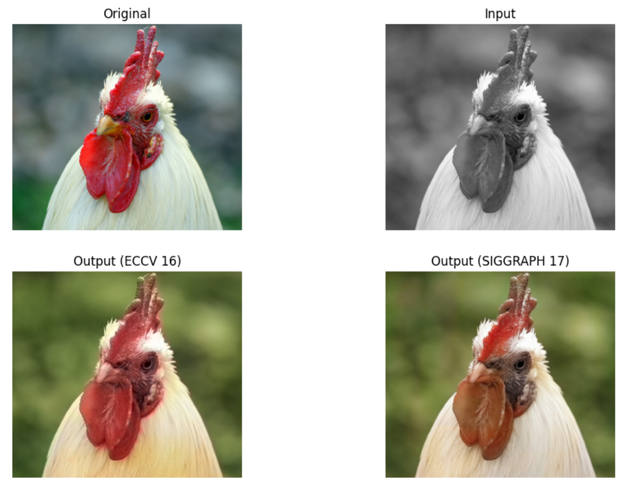

Fig. 2. image 1

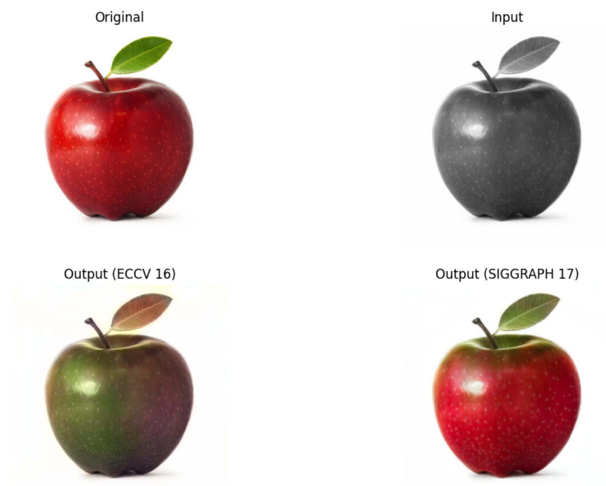

Fig. 3. image 2

The ECCV 2016 method (Fig. 3) produces a reasonable color approximation, with a greenish tint in the background and slightly duller colors on the rooster's comb and wattle. In contrast, the SIGGRAPH 2017 method offers more vibrant and accurate colors, particularly in the red areas of the rooster's comb, making it closer to the original image. While both methods struggle with background accuracy, the SIGGRAPH 2017 result appears more visually appealing and realistic, overall providing a closer match to the original colored image.

In this landscape image(Fig. 4), the ECCV 2016 and SIGGRAPH 2017 colorization methods display distinct stylistic choices. The ECCV 2016 output adds a dramatic, warm-toned overlay, especially in the sky and mountain areas, giving the image a surreal, sunset-like atmosphere with purples and oranges. This stylization, while visually appealing, seems less realistic. The SIGGRAPH 2017 output, on the other hand, produces a more natural look, with realistic greens for the vegetation, blues for the sky, and a more neutral tone for

Fig. 4. image 3

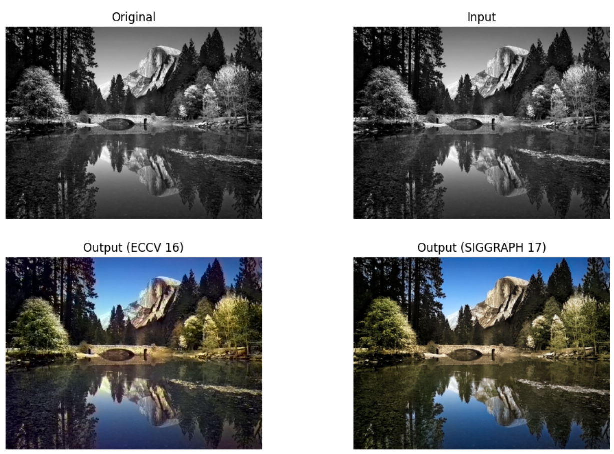

Fig. 5. image 4

the mountains and river. The SIGGRAPH 2017 method better preserves the natural color palette, creating an image that feels more true to life and closer to what the original colors might have been.

VI. EVALUATION

The outputs of the two models ECCV16 and SIGGRAPH 17 are compared using the following matrices;

a) PSNR (Peak Signal-to-Noise Ratio):

This measures the quality of the colorized image by comparing it to a reference image. Higher PSNR values indicate better quality.

b) MSE (Mean Squared Error):

This calculates the average squared difference between the original and the colorized image. Lower MSE values mean the images are more similar.

c) SSIM (Structural Similarity Index):

This measures the similarity between two images based on their structure, luminance, and contrast. SSIM values range from -1 to 1, with 1 indicating perfect similarity. It's a more advanced way to see how similar the images are in terms of their overall structure and appearance.

1. For image 1

SIGGRAPH 17: PSNR of 37.37, MSE of 0.00018, and SSIM of 0.99883.
ECCV 16: PSNR of 35.97, MSE of 0.00025, and SSIM of 0.99805.

The SIGGRAPH 17 output has a higher PSNR, lower MSE, and higher SSIM, indicating better numerical performance. Visually, the SIGGRAPH 17 output is also more vibrant and natural, making it the superior choice in both visual appeal and numerical metrics.

2. For iamge 2

ECCV16 - PSNR: 44.96, MSE: 0.00003, SSIM: 0.99962
SIGGRAPH17 - PSNR: 29.08, MSE: 0.00124, SSIM: 0.97991

The output of SIGGRAPH-17 does indeed seem to be more faithful to the input image. Despite the ECCV16 method having higher PSNR and SSIM values, which indicate better numerical performance, the visual appeal and color accuracy of SIGGRAPH-17 make it stand out. The colors in the SIGGRAPH-17 output are more vibrant and natural, closely resembling what one might expect in a real-life scene. This makes SIGGRAPH-17 a compelling choice for colorization in this case.

3. For image 3

ECCV16: PSNR of 47.92, MSE of 0.00002, SSIM of 0.99919.
SIGGRAPH17: PSNR of 50.15, MSE of 0.00001, SSIM of 0.99987.

The SIGGRAPH17 method outperforms ECCV16 in all numerical metrics, with higher PSNR, lower MSE, and higher SSIM. This indicates that SIGGRAPH17 provides better numerical performance and structural similarity. Visually, the SIGGRAPH17 output is also more vibrant and natural, making it the superior choice for colorization in this case.

4. For image 4

ECCV16: PSNR of 40.95, MSE of 0.00008, SSIM of 0.99852.
SIGGRAPH17: PSNR of 38.52, MSE of 0.00014, SSIM of 0.99898.

The ECCV16 method has higher PSNR and lower MSE, indicating better numerical performance. However, the SIGGRAPH17 method has a slightly higher SSIM, suggesting better structural similarity and visual quality. Despite ECCV16's superior numerical metrics, the SIGGRAPH17 output is often more visually appealing and accurate in terms of color representation.

VII. CONCLUSION

This research aimed to determine a model for automatically colorizing grayscale images using pre-trained models through transfer learning. By leveraging deep learning techniques, particularly CNNs, this paper demonstrated that transfer learning enhances the colorization process, making it more realistic and efficient. Comparing ECCV16 and SIGGRAPH17 models, it can be found that SIGGRAPH17 consistently produced more vibrant and natural colors, making it ideal for applications requiring high visual appeal, such as historical restoration and media. Conversely, ECCV16 achieved higher structural accuracy and better numerical metrics in certain images, making it suitable for applications demanding precise structural fidelity, such as medical imaging. This project highlights the importance of selecting the right model based on application needs to achieve high-quality and contextually appropriate colorizations.

REFERENCES

[1] Liu, Shiguang, and Xiang Zhang. "Automatic grayscale image colorization using histogram regression." Pattern Recognition Letters 33.13 (2012): 1673-1681.

[2] Cheng, Zezhou, Qingxiong Yang, and Bin Sheng. "Deep colorization." Proceedings of the IEEE international conference on computer vision , (2015).

[3] Suárez, Patricia L., Angel D. Sappa, and Boris X. Vintimilla. "Infrared image colorization based on a triplet dcgan architecture." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017.

[4] Deshpande, Aditya, et al. "Learning diverse image colorization." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[5] Varga, Domonkos, and Tamás Szirányi. "Twin deep convolutional neural network for example-based image colorization." Computer Analysis of Images and Patterns: 17th International Conference, CAIP 2017, Ystad, Sweden, August 22-24, 2017, Proceedings, Part I 17. Springer International Publishing, 2017.

[6] Nazeri, Kamyar, Eric Ng, and Mehran Ebrahimi. "Image colorization using generative adversarial networks." Articulated Motion and Deformable Objects: 10th International Conference, AMDO 2018, Palma de Mallorca, Spain, July 12-13, 2018, Proceedings 10. Springer International Publishing, 2018.

[7] Zhao, Jiaojiao, et al. "Pixel-level semantics guided image colorization." arXiv preprint arXiv:1808.01597 (2018).

[8] An, Jiancheng, et al. "Image colorization with convolutional neural networks." 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Ieee, 2019.

[9] Ozbulak, Gokhan. "Image colorization by capsule networks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019.

[10] Su, Jheng-Wei, Hung-Kuo Chu, and Jia-Bin Huang. "Instance-aware image colorization." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[11] Wu, Yanze, et al. "Towards vivid and diverse image colorization with generative color prior." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

[12] Wu, Min, et al. "Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space." The Visual Computer 37 (2021): 1707-1729.

[13] Wang, Yi, et al. "PalGAN: Image colorization with palette generative adversarial networks." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

[14] Wu, Di, et al. "Fine-grained semantic ethnic costume high-resolution image colorization with conditional GAN." International Journal of Intelligent Systems 37.5 (2022): 2952-2968.

[15] Zhang, Jiangning, et al. "Scsnet: An efficient paradigm for learning simultaneously image colorization and super-resolution." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 3. 2022.

[16] Wang, Na, Guo-Dong Chen, and Ying Tian. "Image colorization algorithm based on deep learning." Symmetry 14.11 (2022): 2295.

[17] Treneska, Sandra, et al. "GAN-based image colorization for self-supervised visual feature learning." Sensors 22.4 (2022): 1599.

[18] Kang, Xiaoyang, et al. "Ddcolor: Towards photo-realistic image colorization via dual decoders." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[19] Li, Bin, et al. "Image Colorization using CycleGAN with semantic and spatial rationality." Multimedia Tools and Applications 82.14 (2023): 21641-21655.

[20] Ambadkar, Tanmay, and Jignesh S. Bhatt. "A simple fast resource-efficient deep learning for automatic image colorization." Color and Imaging Conference. Vol. 31. Society for Imaging Science and Technology, 2023.