Image Colorization Using Self-Supervised Learning

Lavanya. Gourish. Nayak[1], Prof. Sandarsh Gowda M M[2]

[1] Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India
[2]Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India

---***---

Abstract

In today's world of data-driven artificial intelligence, the lack of labeled datasets is still a major obstacle for improving computer vision. Self-supervised learning (SSL) has emerged as an effective approach to tackle this issue by using large amounts of unlabeled data through well-designed proxy tasks. This paper explores SSL-based image colorization. By predicting the color channels from a grayscale image, the network learns meaningful and structural representations without needing manual labels. Our method combines convolutional neural networks (CNNs) and generative adversarial networks (GANs) in the Lab color space, which results in vivid, realistic, and context-aware colorizations. By optimizing both reconstruction and adversarial goals together, the model captures detailed textures and overall scene context. Experimental results show high perceptual quality and a strong ability to transfer learned features to other vision tasks. This positions SSL-based colorization as a scalable, cost-effective, and flexible solution for current AI applications.

Keywords—SSL, GAN, Lab color space, Image colorization, Convolutional neural network.

I. INTRODUCTION

Image colorization is the process of adding reasonable color values to a grayscale image. It poses a visually interesting and technically challenging problem in computer vision. The difficulty comes from its inherently undefined nature. A single luminance pattern can relate to several valid color mappings, influenced by object identity, scene context, lighting, and cultural color meanings. This uncertainty makes it hard to rely solely on pixel-level mapping; a deeper understanding of the scene is necessary. By viewing colorization as a self-supervised learning (SSL) task, where the model predicts missing color channels from luminance inputs, we can extract strong and useful visual features without needing manual annotations. This method turns colorization from a purely generative task into an effective way to learn representations, allowing for scalable model development across various fields.

Self-supervised colorization has become a useful technique for using large collections of unlabeled images to learn both basic and complex visual features. Convolutional neural networks (CNNs) are effective at capturing spatial structures, local textures, and edge patterns. Generative adversarial networks (GANs) improve realism, variety, and visual quality by promoting outputs that fit natural image statistics. Using perceptually uniform color spaces like Lab simplifies the problem, splitting luminance from chrominance to lower complexity and speed up learning. This allows the network to focus on producing bright, relevant colors that reflect the relationships in the scene, resulting in outputs that are both visually attractive and semantically accurate.

In this work, we propose an SSL-based image colorization framework that combines CNN-driven feature extraction with GAN-based refinement to find a balance between structural accuracy and visual realism. The method uses a joint optimization of reconstruction loss, which maintains spatial and textural quality, and adversarial loss, which guides the network toward creating more natural and convincing colors. This two-part strategy enables the model to learn features that are effective for color restoration and highly applicable to tasks like classification, segmentation, retrieval, and crowd counting. Experimental results show that this approach achieves performance on par with or better than supervised methods, despite not needing labeled data during training. These findings show that SSL-based colorization is a scalable, efficient, and adaptable technique with the potential to advance generative modeling and other areas of computer vision.

II. LITERATURE SURVEY

Automatic image colorization has progressed along two complementary axes: improving generative quality of the colorized output, and exploiting colorization as a pretext task for representation learning. Early modern approaches reframed colorization as a multimodal prediction problem by casting chrominance prediction into a classification over quantized color bins, thereby encouraging diverse and saturated outputs rather than mean-biased color regression. The seminal “Colorful Image Colorization” work formalized this classification framing and introduced class-rebalancing to promote vivid results. Building on these ideas, methods that operate in perceptually motivated color spaces (e.g., Lab) simplify the learning objective by decoupling luminance and chrominance channels; automatic Lab-space transfer approaches demonstrated the utility of this representation for stable color transfer and reconstruction.

Generative adversarial networks (GANs) and encoder–decoder (U-Net) variants have been widely adopted to improve realism and remove desaturation artifacts. Conditional convolutional GANs trained to predict a and b channels given L have produced sharper and more realistic colorizations than plain L1/L2 losses, aided by architectural elements such as skip connections and perceptual/adversarial losses. More recent generative-prior methods augment the colorization network with a pre-trained generative model

(GAN encoder) to retrieve and inject plausible color priors, enabling vivid and controllable colorizations without external exemplars.

Beyond producing pleasing color images, researchers have exploited colorization as a self-supervised proxy task to learn transferable visual representations. Training networks to predict chrominance from luminance forces models to capture textures, object boundaries, and semantics—features useful for downstream tasks. Empirical evaluations show that colorization pretraining yields feature embeddings competitive with supervised pretraining on several transfer benchmarks. GAN-based colorization used as pretext learning has also been shown to improve downstream classification and segmentation performance by encouraging perceptually coherent feature learning.

Hybrid and task-aware SSL designs extend single-task colorization by combining it with complementary pretext objectives to strengthen generalization and domain adaptation. Methods that fuse global priors or cluster-level signals with local colorization learning have been employed for specialized applications such as crowd counting, where merged global priors, semantics, and texture features reduce reliance on labeled density maps and close the gap to supervised baselines using only limited annotations. More generally, multi-task SSL frameworks that combine colorization with spatial reasoning, relational objectives, or other proxy tasks improve robustness across domains and are particularly helpful when transferring to new visual domains.

A second important strand of literature focuses on designing more effective SSL algorithms and training regimes. Works advocating relational or distributional consistency show that modelling inter-instance relations (rather than only instance discrimination) can boost representation quality and training efficiency. Other studies emphasize scaling strategies and resource efficiency — for example, showing that smaller input resolutions and partial backbones can produce competitive SSL representations with dramatically lower compute costs, a finding that is relevant when pretraining colorization models on limited compute or domain-specific datasets. Complementary SSL research explores alternative pretext strategies (artifact spotting, multi-instance contrastive schemes) that provide different inductive biases; such methods can be combined with colorization to enrich the learned features.

Specialized applications and domain adaptations of SSL colorization demonstrate the method's flexibility. In biomedical imaging, pseudo-colorization of masked cells (with physics-informed colormaps) served as a domain-tailored pretext task that implicitly learned segmentation-like features and outperformed several generic SSL baselines on microscopy tasks. For geometric and local descriptor learning, improvements in sampling and hard negative mining yield stronger local features even when supervision is only self-generated, showing that SSL improvements are not limited to global classification tasks. Active learning pipelines have also been proposed where simple pretext task losses (including colorization) guide selection of samples for annotation, improving overall annotation efficiency.

Synthesis & Gap. Taken together, the literature shows that colorization is a powerful SSL proxy that simultaneously addresses generative quality and representation learning. Generative models (GANs and generative priors) improve perceptual realism, while classification-style formulations and Lab-space modeling stabilize training and diversity. Hybrid SSL strategies and relational/contrastive advances further strengthen transferability and domain adaptation. Remaining gaps include principled methods that combine generative color priors with relational SSL objectives to maximize both visual fidelity and representation robustness; efficient SSL recipes tailored to low-compute or small-data domains; and standardized benchmarks that jointly evaluate colorization fidelity and downstream task transfer under the same pretraining regime. Our work builds upon these directions by integrating adversarial refinement, Lab-space modeling, and SSL training choices that target both perceptual quality and downstream transferability.

III. EXISTING SYSTEM

Current image colorization methods mainly use supervised deep learning techniques, which need a lot of paired grayscale and color images for training. These methods often use convolutional neural networks (CNNs) and generative adversarial networks (GANs). They work in color spaces like Lab to predict the color information from the brightness information. Improvements like skip connections, quantized color classification, and perceptual loss functions have made the outputs more realistic and varied, helping models create appealing results.

However, these supervised systems depend heavily on labeled datasets, making them costly to develop and tough to adjust for areas where color references are not available. Their performance often drops when applied to images from different domains not included in the training set, which limits their ability to generalize. Furthermore, existing methods mainly focus on visual quality and do not take advantage of colorization as a self-supervised learning (SSL) tool to learn strong, transferable features for tasks such as classification, segmentation, and retrieval.

Disadvantages

- Requires large amounts of labeled training data, increasing annotation cost.
- Poor generalization to unseen domains or datasets.
- High computational cost for training on large-scale datasets.
- Focuses mainly on visual quality, not on learning transferable representations.
- Limited adaptability for low-data or domain-specific scenarios.

IV. PROPOSED SYSTEM

The proposed system uses self-supervised learning (SSL) for image colorization without requiring large labeled datasets. It predicts the chrominance (a and b) channels from the luminance (L) channel in the Lab color space. This helps the model learn meaningful semantic and structural features from unlabeled images. A convolutional neural network (CNN) captures spatial and texture information, while a generative adversarial network (GAN) improves outputs to ensure bright, realistic, and context-aware colors. The framework optimizes both reconstruction and adversarial losses, balancing spatial accuracy with perceptual quality. This design allows the model to generalize well across various domains and supports transferable feature learning for tasks like classification, segmentation, and retrieval. It overcomes the limitations of traditional supervised colorization methods.

Advantages:

- A thorough and efficient bug prediction model is produced by combining ML, NLP, and text-mining techniques.
- Eliminates dependency on large labeled datasets, reducing annotation costs.
- Produces vivid and realistic colors with high perceptual quality.
- Learns transferable features useful for multiple downstream vision tasks.
- Robust to domain shifts, improving generalization.
- Scalable and efficient for large-scale unlabeled datasets.

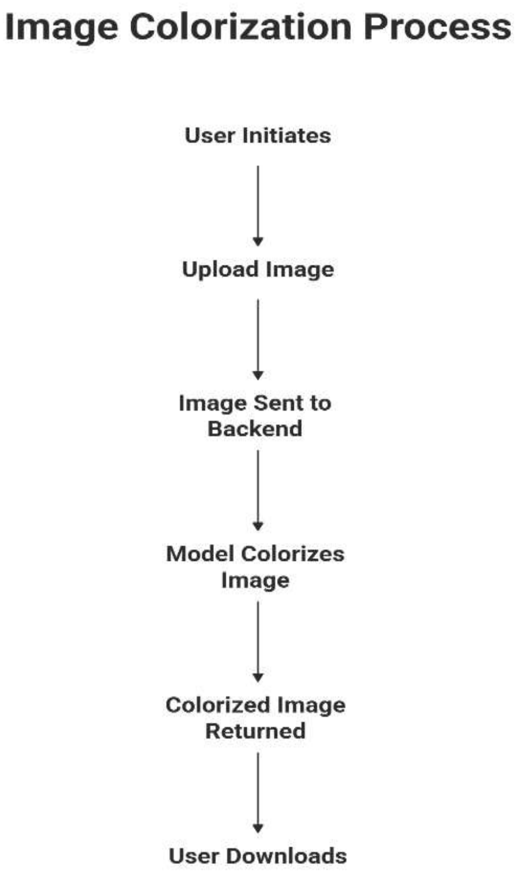

Fig 1: Proposed Model

V. IMPLEMENTATIONS

Data Preparation:
The dataset was obtained from Kaggle, consisting of unlabeled images. Each image was converted into the Lab color space, with the L channel as input and the a, b channels as targets. Preprocessing included resizing, normalization, and augmentation (rotation, flipping, cropping) to improve generalization.

Model Development
The system was implemented in Python using TensorFlow, Keras, and PyTorch. A CNN backbone extracted low-level and high-level features. A GAN framework with a generator and discriminator improved perceptual realism. A Hugging Face pre-trained model was used for transfer learning to reduce training time and improve convergence.

Training Strategy:
Training used a self-supervised approach. It took the luminance channel as input and treated the color channels as pseudo-labels. The goal combined reconstruction loss (L1), adversarial loss, and perceptual loss. We optimized using Adam with learning rate scheduling and early stopping. Training was sped up on Google Colab GPUs that supported CUDA.

System Deployment:
The application was set up as a web-based system with a React.js frontend for uploading images and displaying results. A Python backend, using Flask or Django, handles model execution. Outputs and metadata are stored in MongoDB. The system provides users with real-time colorization results.

Performance Evaluation:
Performance was evaluated using PSNR and SSIM for numerical assessment, along with human inspection for qualitative insights. The SSL-based model produced bright, context-rich colors. It achieved results similar to supervised methods while lessening the need for labeled data.

Transformation:
Next, these tokens are converted into numerical forms so that machine learning processing can occur.

VI. CONCLUSIONS

In this study, we implemented an Image Colorization system using Self-Supervised Learning (SSL) to reduce the reliance on large labeled datasets and improve the consistency of grayscale-to-color transformation. By using the Lab color space, the model predicted chrominance channels from luminance input. This approach allowed it to learn semantic and structural features without needing manual annotations. We employed a CNN-GAN hybrid framework, where the CNN extracted multi-level representations, and the GAN improved perceptual realism. Transfer learning from Hugging Face models also enhanced convergence efficiency. The training process combined reconstruction, adversarial, and perceptual losses, optimizing with Adam and GPU acceleration to ensure balanced performance. We deployed the system as a web-based application, integrating a React.js frontend, a Python backend, and MongoDB for real-time and user-friendly interaction. Evaluation through PSNR, SSIM, and qualitative analysis showed vivid and contextually accurate outputs. This confirms that SSL-based colorization offers a scalable, efficient, and annotation-free method for both image restoration and broader vision applications.

VII. FUTURE ENHANCEMENTS

In the future, we can extend the proposed system by adding Vision Transformers (ViT) and self-supervised frameworks. These additions will help capture long-range dependencies and improve contextual color predictions. Combining multi-modal learning with text prompts or scene descriptions could allow for controllable and user-guided colorization. Expanding the dataset with domain-specific images, such as medical, satellite, or historical archives, would improve generalization in specialized applications. We can optimize the deployment pipeline with lightweight models and edge computing support to enable real-time performance on mobile and embedded devices. Additionally, using advanced evaluation metrics, like Fréchet Inception Distance (FID), may give us better insights into perceptual quality. Together, these improvements aim to boost scalability, flexibility, and usability, making the system a strong tool for both academic research and real-world applications.

VIII. REFERENCES

[1] H. Bai, S. Wen, and S.-H. G. Chan, “Crowd Counting by Self-supervised Transfer Colorization Learning and Global Prior Classification,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020.

[2] S. Bucci, A. D’Innocente, Y. Liao, F. M. Carlucci, B. Caputo, and T. Tommasi, “Self-Supervised Learning Across Domains,” in *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4756–4769, Sep. 2022.

[3] Y.-H. Cao and J. Wu, “Rethinking Self-Supervised Learning: Small is Beautiful,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021.

[4] N. K. El Abbadi and E. Saleem, “Automatic Gray Images Colorization Based on Lab Color Space,” *Int. J. Comput. Appl.*, vol. 176, no. 26, pp. 22–27, Jul. 2020.

[5] I. Melekhov, Z. Laskar, X. Li, S. Wang, and J. Kannala, “Digging Into Self-Supervised Learning of Feature Descriptors,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019.

[6] K. Nazeri, E. Ng, and M. Ebrahimi, “Image Colorization Using Generative Adversarial Networks,” in *Proc. Int. Conf. Artif. Intell. Data Process. (IDAP)*, 2018.

[7] V. Rani, S. T. Nabi, M. Kumar, A. Mittal, and K. Kumar, “Self-Supervised Learning: A Succinct Review,” *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 9, pp. 6736–6752, Sep. 2023.

[8] S. Treneska, E. Zdravevski, I. M. Pires, P. Lameski, and S. Gievska, “GAN-Based Image Colorization for Self-Supervised Visual Feature Learning,” in *Proc. Int. Conf. Bioinformatics Biomedicine (BIBM)*, 2022.

[9] R. Wagner, C. F. Lopez, and C. Stiller, “Self-Supervised Pseudo-Colorizing of Masked Cells,” in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, 2022.

[10] Y. Wu, X. Wang, Y. Li, H. Zhang, X. Zhao, and Y. Shan, “Towards Vivid and Diverse Image Colorization with Generative Color Prior,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021.

[11] J. S. K. Yi, M. Seo, J. Park, and D.-G. Choi, “PT4AL: Using Self-Supervised Pretext Tasks for Active Learning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[12] R. Zhang, P. Isola, and A. A. Efros, “Colorful Image Colorization,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 649–666.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.

[14] M. Zheng, S. You, F. Wang, C. Qian, C. Zhang, X. Wang and C. Xu, “ReSSL: Relational Self-Supervised Learning with Weak Augmentation,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2021.

[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020.