# Generative AI for Cardiac Organoid Florescence Generation

Arun Kumar Reddy Kandula[1], Tanakit Phamornratanakun[1], Angello Huerta Gomez,[2] Rahulkumar Bhoi,[2] Marcel El-Mokahal,[2] Yunhe Feng[1*], Huaxiao Yang[2*]

[1] Department of Computer Science & Engineering, University of North Texas

[2] Department of Biomedical Engineering, University of North Texas

*: Co-corresponding author, Yunhe Feng, Yunhe.Feng@unt.edu, and Huaxiao Yang, Huaxiao.yang@unt.edu

## ABSTRACT

hPSC-derived cardiac organoid is the most recent three-dimensional tissue structure that mimics the structure and functionality of the human heart and plays a pivotal role in modeling heart development and disease. The hPSC-derived cardiac organoids are commonly characterized by bright-field microscopic imaging. Although the bright-field microscope provides essential information about hPSC-derived cardiac organoids, such as morphology, size, and general structure, it does not extend our understanding of cardiac organoids on cell type-specific distribution and structure. Then, fluorescence microscopic imaging is required to identify the specific cardiovascular cell types in the hPSC-derived cardiac organoids by fluorescence immunostaining of fixed organoid samples or fluorescence reporter imaging of live organoids. Both approaches require extra steps of experiments and do not provide general information on all hPSC-derived cardiac organoids from different hPSC lines, which limits the biomedical applications of hPSC-derived cardiac organoids. This research addresses this limitation by proposing a comprehensive workflow for colorizing grayscale cardiac organoid images from bright-field microscopic imaging using conditional Generative Adversarial Networks (GANs) to provide cell type-specific information. By

infusing these grayscale images with accurate fluorescence colorization, our approach aims to unlock the hidden wealth of cell type, distribution, and network, for bettering characterizing hPSC-derived cardiac organoids.

## INTRODUCTION

Various human organ-specific organoids, such as hepatic organoids, cerebral organoids, intestinal organoids, kidney organoids, and most recently, cardiac organoids, have been developed and employed in drug modeling and drug discovery. To comprehend organoid development and its response to specific drugs, fluorescence microscopic imaging is essential for observing the fluorescence emitted by these cells. Cardiac organoids typically emit three different fluorescence signals: Blue (Cyan Fluorescent Protein - CFP) representing smooth muscle cells, green (Green Fluorescent Protein - GFP) representing cardiomyocytes, and red/orange (Red Fluorescent Protein/Monomeric Orange Protein - RFP/mOr) representing endothelial cells. Since each fluorescence signal corresponds to a specific cell type, generating accurate fluorescence information from the phase images of organoids is crucial for a precise analysis of cardiac organoid development.

Numerous approaches have been explored to tackle the challenge of image colorization. Traditional methodologies rely on machine learning techniques that extract similar features from a reference image to predict colors in a new image [3, 6, 27]. Notably, the efficacy of such methods is contingent on the similarity between the reference image and the target, particularly in terms of cell type. The advent of Convolutional Neural Networks (CNNs) marked a significant shift, allowing for the automatic extraction of features from images. Pretrained CNNs have gained prominence in image colorization, leveraging feature maps to predict pixel colors [7, 12, 26]. The capabilities of Generative Adversarial Networks (GANs) [4] in various generative tasks have prompted their use in colorization. In this context,

Conditional GANs, exemplified by Pix2Pix GAN [8], have emerged, mapping grayscale inputs to corresponding ground truth images. In our work, we employ Pix2Pix GAN, augmented with the Convolutional Block Attention Module (CBAM) [24], enhancing the network's focus on critical features and elevating colorization realism.

Most existing research and methods in image colorization are geared towards generic image categories, often trained on datasets like ImageNet and COCO. Notably, there is a lack of research focused specifically on colorizing cardiac organoids. While small color discrepancies in generated images might be tolerable for generic categories, the same cannot be said for cardiac organoids. Even minor color variations in this context can introduce significant misinformation, rendering the task of organoid colorization exceptionally challenging.

Our work introduces a novel solution to the challenging problem of colorizing phase images of cardiac organoids. Using a Conditional Generative Adversarial Network (cGAN), with an iterative interaction between the generator and a discriminator. This GAN-based model learns to capture and replicate the intricate color information present in the cells. To address the dynamic nature of cardiac organoid images, we incorporated an attention mechanism, the Convolutional Block Attention Module (CBAM), ensuring an increased emphasis on crucial details. Through an extensive training process on our dataset, the model learns to intricately map grayscale cardiac Organoid images to the corresponding color images. This solution not only tackles the inherent challenges of generic colorization methods but also emerges as a specialized tool tailored to the distinctive characteristics of cardiac organoid imaging. The proposed framework demonstrates its efficacy in preserving cell-level information, presenting a promising advancement for the visualization and analysis of organoid structures in biomedical research.

This research makes several contributions to the field of Organoid image colorization: We proposed a novel framework utilizing GANs for training on cardiac organoid images. The integration of the

Convolutional Block Attention Module (CBAM) enhances the model's capacity to focus on critical cell-level features, capturing fine details and generating more accurate colors. We conducted a thorough evaluation of the proposed method's effectiveness in preserving biological details and introduced a new evaluation metric, the Weighted Patch Histogram, designed to capture the color histogram information from small patches of the image, allowing us to obtain a spatially aware color histogram. A comparison of these spatially aware histograms is a more suitable comparison compared to the widely used metrics in image colorization problems like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) because they are not specifically designed to measure color similarity.

## RELATED WORK

This section will provide an overview of the related image colorization works. We categorize them into three categories based on their underlying techniques.

**Reference Image-based colorization:** In this technique the target image is colorized based on the color information from a reference image. Gupta et al. [5] colorized the target images using a reference-colored image, where feature mapping is done for the features extracted using SURF and Gabour filters, and image space voting based on the neighboring pixels is done to obtain the plausible pixel color. This technique suffered at image boundaries and caused color bleeding. To solve this problem, a patch-based feature extraction and colorization technique was proposed in [11], that produced more robust colors. Similarly, Fang et al. [3] also proposed a technique using superpixel, but the color assignment to a pixel is done by determining a candidate set of the possible colors, then the variational energy factors that enforce spatial consistency and non-local self-similarity constraints, which helped determine the most possible pixel color. This reference image-based technique was applied in the biomedical domain by [27]. With the breakthrough of Convolutional Neural Networks (CNN) performance in Image processing tasks, recently a

lot of research in image colorization was based on CNN. He et al. [6], used VGG-19 to generate feature maps based on the grayscale target and reference image. These feature maps were combined with the grayscale target image and the color information in terms ab color channels, from CIE Lab color space. This combined input was passed to an encoder-decoder network to generate the final colored image.

**CNN-based image colorization:** CNNs are being used widely because of their capability to automatically extract the features to find the relations between them and produce more realistic colors. Larsson et al. [12] used a VGG-16 architecture, in which the first convolution layer was modified to operate on a single channel, and the classification layer was removed. This architecture was fine-tuned on the ImageNet dataset of gray images for one epoch. The grayscale images were passed to this modified network to generate spatially localized multicolumn layers referred to as hypercolumns, which were used to predict the color of the pixel. Iizuka et al. [7] used multiple CNN architectures, one to extract global level features of the image and the other to extract mid-level features, and created a fusion layer that combined these two global and mid-level features now the final color predictions were generated based on the correlation of those two features. Zhang et al. [26] also used an off-the-shelf VGG network and modified the loss functions. They used classification loss, with rebalanced rare classes. In this network, they operated in CIE Lab colorspace to generate a color mapping function of the corresponding (ab) channels from the input Lightness (L) channel. This work is considered to produce state-of-the-art results. Similarly, there are many implementations of image colorization based on Vgg-16 with modifications in their operating color space and the loss functions like [1, 13, 21].

**Generative Adversarial Network (GAN) based image colorization:** Pix2Pix [8], a type of Conditional GAN is becoming a popular choice for image colorization because of their ability to find information in pair-to-pair image translation. GAN network consists of a generator, which generates the colorized image from a conditional input image, and a discriminator, which tries to identify if the generated image is real or fake. Suarez et al. [19], used GAN architecture to generate´ colored images from Infrared images. They are

operating in RGB color space and using 3 generator networks, which were given a conditional input of the infrared image. Each generator generated a corresponding color channel of RGB, and the combined RGB image was passed to the discriminator network to check the probability of the generated image being real. Nazeri et al. [14], used pix2pix architecture for image colorization on the CIFIAR-10 dataset, where the grayscale image was given as the conditional input to the U-Net generator to generate a colored image, which was then passed to the discriminator to identify if the generated image is a real or fake one. Treneska et al. [20], is also implemented a pix2pix architecture, where the grayscale image was given as the conditional input to the U-Net generator architecture to generate a colored image, and passed it to a patch discriminator to determine if the generated image was real or fake. The patch discriminator divided the image into NxN patches to produce the probability of a patch being real or fake and averaged the results of all patches to determine if the generated image is real or fake, unlike a regular discriminator that used the whole image at once to make the prediction. Kiani et al. [10], also used a similar approach with pix2pix for image colorization.

In our research, we are also exploiting the strengths of image-to-image translation of Pix2pix architecture. To make the model more robust to pay more attention to the details we included Convolution Block Attention Module (CBAM) [24] in the network.

## METHODOLOGY

### A. Framework Overview

In this section, we present an overview of our research on image colorization of grayscale cardiac organoids using conditional Generative Adversarial Networks (GAN), specifically the Pix2Pix model [8]. Image colorization is an essential task in medical image processing, enabling a more comprehensive and intuitive visualization of grayscale images. Grayscale cardiac organoid images, acquired through various

medical imaging techniques, lack color information, making it challenging for medical professionals to interpret and analyze them effectively. Our research aims to address this limitation by employing conditional GANs, which have shown promising results in generating realistic and accurate colorizations from grayscale images. In this paper, we present a comprehensive workflow for cardiac organoid image colorization, highlighting the motivations behind using conditional GANs for this specific task.

Our methodology is built around utilizing the Pix2Pix conditional GAN [8], the Pix2Pix model, short for ”Pixel-toPixel Translation,” is a notable example of a conditional Generative Adversarial Network (GAN). GANs are a class of deep learning architectures that involve two networks, a generator, and a discriminator, engaged in a competitive learning process. In the context of image colorization, the Pix2Pix model learns to map input grayscale images to corresponding colorized versions by leveraging a training dataset that contains pairs of grayscale and color images. By doing so, the model can generate plausible and realistic colorizations that align with the original context of the grayscale input. Pix2Pix has garnered attention due to its ability to capture intricate relationships between input and output images, making it a compelling choice for various image translation tasks, including our focus on cardiac organoid image colorization.

This robust pix2pix deep-learning framework is proficient in mapping input grayscale images to their corresponding colorized versions. To achieve this, we adopt the CIELAB color space, consisting of three channels: Lightness, a*, and b* [as per figure]. In CIELAB, Lightness represents the grayscale channel, while a* and b* represent the two-color channels. This Lightness channel will serve as the conditional input to the generator, and the a* and b* channels will be the target channels for generating colorized versions of the grayscale images. The objective of using CIELAB color space is to extract only the color information from the cardiac organoid and train the model to generate the plausible colors of a* & b* that will be merged on the grayscale input, to obtain the colorized cardiac organoid.

Additionally, we have incorporated the Convolution Block Attention Module (CBAM) [24] to increase the channel and spatial attention of the GAN model to focus on the relevant features. CBAM is an innovative enhancement introduced to the architecture of deep neural networks, particularly Convolutional Neural Networks (CNNs). CBAM integrates both channel and spatial attention mechanisms, facilitating the model's ability to focus on pertinent features within the input data. Channel attention enables the network to adaptively assign importance to different channels, emphasizing relevant information while suppressing noise. Simultaneously, spatial attention ensures that the network allocates its focus to meaningful spatial regions within an image. By integrating CBAM into our conditional GAN framework, we aim to improve the model's ability to capture salient features in grayscale cardiac organoid images. This attention mechanism enhances the quality and fidelity of the generated colorizations by directing the model's focus toward critical regions within the image.

The primary motivation for incorporating conditional GANs, CIELAB color space, and CBAM is to increase the model's attention to relevant features and limit the model's predictions to only two channels (i.e a* & b*), thereby reducing the number of predictions compared to the RGB color space, where the model would have to make predictions for the R, G, B channels. The synergy between Pix2Pix, CIELAB, and CBAM contributes to notable colorization outcomes.

To provide a visual representation of our image colorization workflow, we present an overview figure (Figure 1) that illustrates the main components and steps of the process. The figure depicts the transformation of a grayscale cardiac organoid image to a fully colorized output using Pix2Pix conditional GAN. The conditional input passed to the Generator is the Lightness channel and the Discriminator is trained on the a* & b* channels.

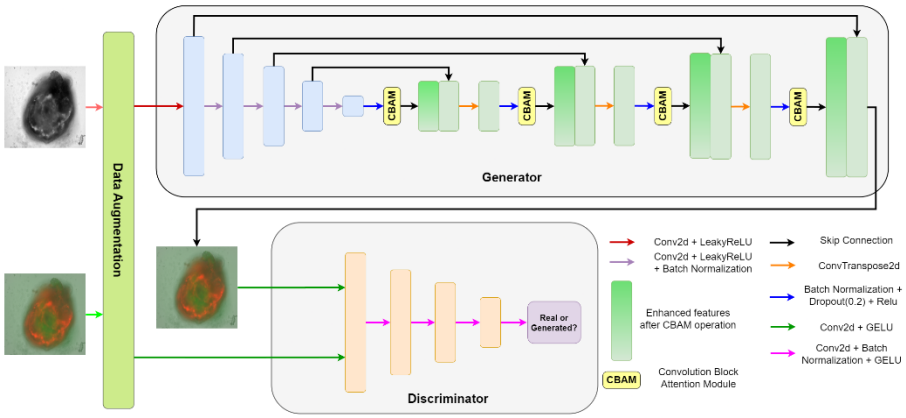



Fig. 1: GAN Architecture Overview

B. Individual Models

*U-Net generator*

The U-Net generator is a fundamental component of our Conditional GAN-based image colorization system. It plays a pivotal role in transforming grayscale organoid images into their corresponding colorized versions. Here, we provide an overview of our U-Net generator's architecture.

The U-Net generator derives its name from its characteristic U-shaped architecture. It consists of an encoder and a decoder, connected by a bottleneck layer. Figure 2 shows the architecture of our U-Net generator where the encoder progressively reduces the spatial dimensions of the input grayscale image while extracting features. The decoder then upsamples these features to produce the final colorized output. Skip connections between corresponding encoder and decoder layers facilitate the flow of low-level features, enhancing the network's ability to capture fine details.

One distinctive feature of our U-Net generator is its utilization of the Lightness (L) channel from the CIELAB color space as a conditional input. This L channel represents the grayscale information of the input image. By incorporating this channel, the generator can focus on producing color information (a*

and b* channels) that is coherent with the grayscale content.

Our research enhances the Generator's ability to focus on relevant features using the Convolution Block Attention Module (CBAM). CBAM integrates channel and spatial attention mechanisms, enabling the discriminator to adaptively assign importance to different channels and meaningful spatial regions within the image.

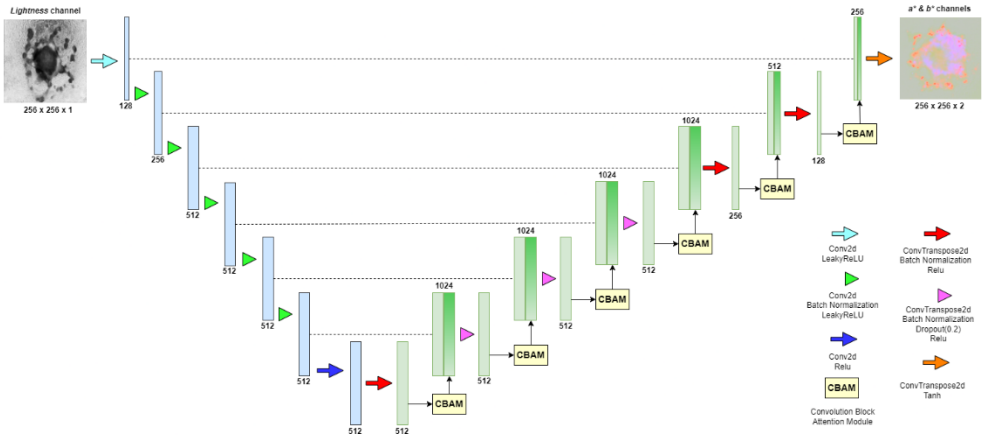



Fig. 2: U-Net generator Architecture with CBAM

*Convolutional Block Attention Module (CBAM)*

CBAM is an integral component incorporated into our UNet Generator architecture to enhance its ability to capture and emphasize relevant features within grayscale organoid images. The input feature map is and CBAM extracts 1D channel attention map and a 2D spatial attention map . In summary, the overall attention processes can be explained as:

(1)

where ⊗ denotes the element-wise multiplication and the resulting F" is the final refined output map that includes the details from both channel attention and spatial attention. This operation allows the model to focus on relevant features while suppressing irrelevant information. A more detailed explanation of the

individual attention maps is available in [24].

**Channel attention** enables the network to adaptively assign importance to different channels of feature maps, emphasizing relevant information while suppressing noise. Channel attention is essential when dealing with multi-channel images such as the L*a*b* color space we operate in. This selective channel weighting allows the model to focus on the most informative colorization components.

**Spatial attention** is another crucial aspect of CBAM. It ensures that the network allocates its focus to meaningful spatial regions within an image. In the context of colorization, this is especially important as it guides the model to concentrate on the relevant regions where colorization details are essential. Spatial attention complements channel attention by pinpointing critical areas in the input.

By integrating CBAM into our conditional GAN framework, we aim to significantly improve the model's ability to capture salient features in grayscale cardiac organoid images. This attention mechanism enhances the quality and fidelity of the generated colorizations by directing the model's focus toward critical regions within the image and generating realistic and accurate colorizations of grayscale organoid images.

*Patch Discriminator*

The Patch Discriminator is a convolutional neural network (CNN) designed to operate on image patches rather than entire images as shown in Figure 3. This approach allows the discriminator to focus on local details and textures, making it well-suited for assessing the quality of colorizations at a fine-grained level. It consists of multiple convolutional layers to produce a single feature map that is used to classify the patch as real or fake. The final classification result for the entire image is obtained by averaging the predictions from the patches across the entire image. The result is a global classification score that represents the discriminator's assessment of the overall image

The Patch Discriminator engages in adversarial training with the U-Net generator. It aims to distinguish between real colorized organoid patches and fake patches generated by the generator. Through this

adversarial process, the discriminator provides feedback to the generator, encouraging it to produce colorizations that are indistinguishable from real color images.

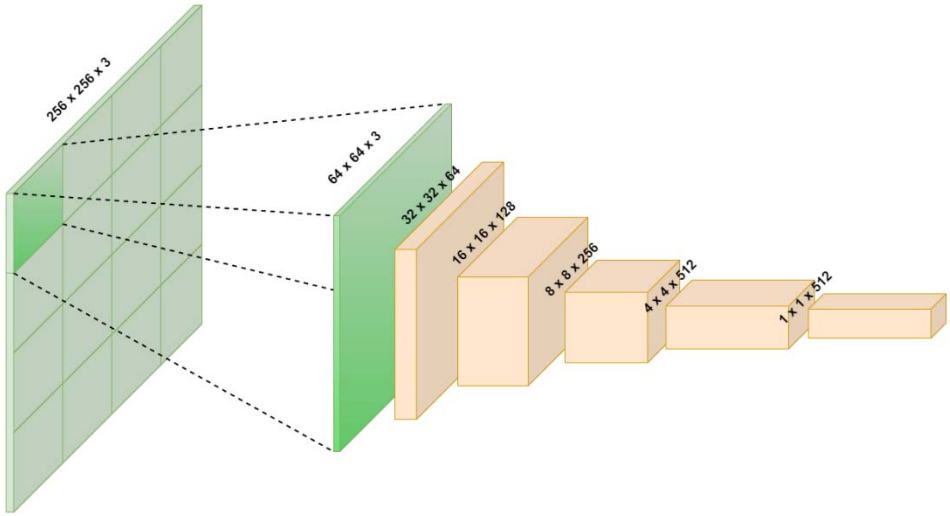



Fig. 3: Patch Discriminator Architecture

The primary objective of the Patch Discriminator is to guide the U-Net generator in generating high-quality colorizations. Assessing the local realism of colorized patches, helps ensure that fine-grained details and textures are faithfully preserved in the output.

*Loss Functions*

**Discriminator Loss:** The Discriminator, a key component of our conditional GAN, serves the crucial role of assessing the authenticity of colorized organoid images. To fulfill this role, we employ the Binary Cross-Entropy Loss (BCEWithLogitsLoss) as our discriminator loss function.

Mathematically, the discriminator loss can be expressed as:

– (2)

Here, represents the discriminator loss, where is the batch size, denotes the ground truth colorized organoid images, represents labels for real images and fake images signifies

the discriminator's output for real images, and $G(L_i)$ signifies the generator's output for the corresponding grayscale input $(L_i)$. The BCEWithLogitsLoss computes the binary cross-entropy loss by comparing the discriminator's predictions with the ground truth labels.

The discriminator aims to maximize this loss, which encourages it to correctly classify real and fake patches within the images. Simultaneously, the generator minimizes this loss during adversarial training to produce colorizations that are indistinguishable from real images.

**Generator Loss:** The Generator, a pivotal component of our conditional GAN, is tasked with generating plausible colorizations. To achieve this, we employ a combination of two loss functions: Binary Cross-Entropy Loss (BCEWithLogitsLoss) and L1 Loss (Mean Absolute Error). Similar to the discriminator, the generator uses BCEWithLogitsLoss as its adversarial loss function. It encourages the generator to produce colorizations that convincingly fool the discriminator into classifying them as real.

Mathematically, the generator's adversarial loss is defined as:

$$\mathcal{L}_{\mathcal{G}_{\mathcal{GAN}}} = -\frac{1}{N}\sum_{i=1}^{N} \log\left(D\left(G(L_i)\right)\right) \tag{3}$$

This loss drives the generator to produce colorizations that are perceptually similar to real color images.

L1 Loss (Mean Absolute Error): In addition to the adversarial loss, we incorporate the L1 Loss to ensure that the generated colorizations closely match the ground truth images in terms of pixel-wise similarity.

Mathematically, the generator's L1 loss is expressed as:

$$\mathcal{L}_{\mathcal{G}_{\mathcal{L}1}} = \frac{1}{N}\sum_{i=1}^{N} |\, G(L_i) - C_i |_1 \tag{4}$$

Here, $L_{G_{L1}}$ represents the generator's L1 loss, where $N$ is the batch size, $L_i$ denotes the grayscale input images, $G(L_i)$ represents the generator's colorized output, and $C_i$ denotes the corresponding ground truth

color images. The L1 loss encourages the generator to produce colorizations that closely match the ground truth, focusing on fine-grained pixel-level details.

By combining these two loss components, the generator is trained to produce colorized organoid images that are both visually convincing and pixel-wise accurate, ultimately enhancing the quality and realism of the generated colorizations

## EVALUATION AND FINDINGS

### *A. Image Similarity Measurement Metrics*

Evaluating the accuracy and quality of the generated image is a challenging task and on top of that, we have a limited dataset of 1300 images so we are using non deep-learning metrics to obtain a similarity score. We are using 3 evaluation metrics PSNR, SSIM, and Weighted Patch Histogram.

Peak Signal-to-Noise Ratio (PSNR), is used in image processing to measure the quality of reconstructed or compressed images and this metric is used widely in many research papers [9, 16, 18, 22, 25] for comparing the similarity of the colorized image with ground truth. It objectively measures how well a colorization technique preserves the details and visual fidelity of the original image. By calculating the PSNR value, we can evaluate the accuracy and fidelity of colorization algorithms. Its range is (0, ∞), 0 represents no similarity between images, and the higher the score higher the similarity.

PSNR score of an *mxn* image *I* and its compressed image *K* can be determined by:

$$20 \cdot \log_{10}(\mathrm{MAX}_I) - 10 \cdot \log_{10}(\mathrm{MSE}) \tag{5}$$

Where $MAX_I$ is the maximum possible pixel value of the image and MSE is the mean square error of the Original Image I and its compressed image K, it can be calculated by:

$$\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2 \tag{6}$$

The **Structural Similarity Index (SSIM)**, is also a widely used metric in research papers [9, 16, 18, 22, 25] for assessing the visual quality of the colorized image with ground truth. It considers global and local image characteristics, capturing the perceptual differences and structural similarities between the colorized and ground truth images. To be specific, SSIM compares three components in an image pair, suppose x and y are the two patches of the true and compressed image respectively that are aligned with each other, the luminescence comparison function l(x,y) captures the differences in brightness, the contrast comparison function c(x,y) accesses variation in image contrast and the structure comparison function s(x,y) measures differences in image structure and texture. SSIM is a combination of all these three factors.

$$SSIM(x,y) = [l(x,y)\cdot c(x,y)\cdot s(x,y)] \tag{3}$$

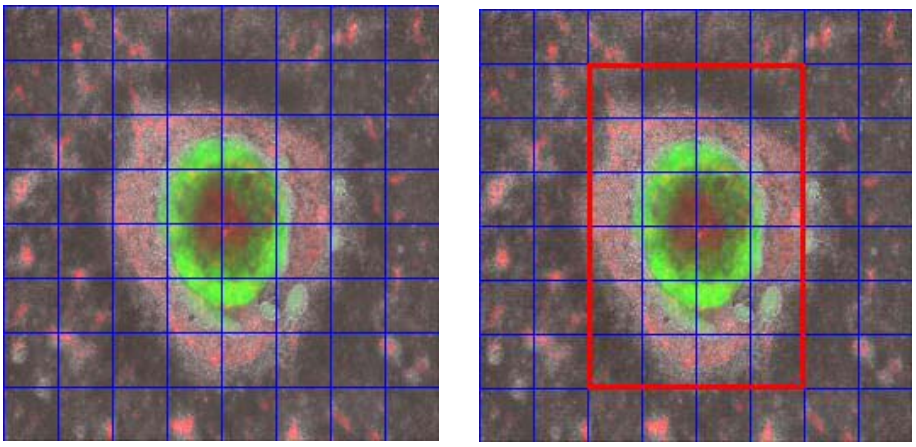
A detailed explanation of these components is in [23]. By evaluating the preservation of underlying structures and textures, SSIM provides a comprehensive measure of the algorithm's ability to maintain visual coherence and realism. The SSIM score typically falls within the range of (-1, 1] [15], where a higher score signifies greater similarity.

PSNR and SSIM are widely used metrics in evaluating Image Colorization tasks. Still, they are not exactly appropriate for the problem, because PSNR is designed to identify the quality of the compressed image with the original image. Similarly, SSIM primarily focuses on structural similarity rather than color which is the main part of Image colorization. So, we are proposing Weighted Patch Histogram to compare the similarity of generated colors. With regular Histogram comparison, we will lose valuable spatial information of the color, so in our approach, we are splitting the image into a 16x16 grid (Figure 4(a)) to have multiple small patches of the image and comparing these small patches individually to the corresponding patch from the ground truth. This patch histogram comparison will increase the spatial

information of the pixel's value.

As patch histogram comparison will increase spatial color information, it is ideal to consider that reducing the patch size to the smallest possible value will produce the best results. The smallest size possible to compare is 1x1 pixels, which will lead us to a pixel-to-pixel comparison of the images and it would be highly sensitive to noise and unreliable. So, we must find the right balance between the patch size and the number of bins in the histogram comparison. After many experiments, we found a patch size of 32x32 pixels and 32 bins ideal for histogram comparison in our scenario. In our dataset, most of the Organoids are in the center of the image and it is our Region of Interest (ROI), so the color comparison is more significant in our ROI rather than the background, so we are increasing the weightage for the patches inside the ROI in Figure 4(b) by 50% to give more importance to the colors in the organoids.

In our research, the size of the images is 256x256, which we are breaking into a 16x16 grid (Figure 4(a)) to give us multiple patches of size 32x23 pixels. Now these patches of the generated image are compared with the corresponding patches of the ground truth in CIELAB color space as the GAN models generated color in CIELAB space.



(a) (b)

Fig. 4: (a) Image divided into 8x8 grid of small patches, (b) Highlighted in red is the ROI, which is given more weightage for histogram comparison.

*B. Basic Results*

In this section, we will provide a detailed analysis of the results obtained with different training strategies that we employed for comparison of the Image colorization on Cardiac Organoids.

*Model 1, U-Net generator:* This generator architecture is similar to Figure 2, excluding the CBAM block and the enhanced output from it.

*Model 2, U-Net generator with CBAM:* This is the same architecture that we discussed in detail in Section III.

*Model 3, U-Net with CBAM and Generator Iteration:* This is a training technique employed on the architecture in Model 2, where the generator is trained twice in an epoch. With an intuition of making the generator stronger compared to the discriminator as it is trained multiple times to produce more realistic colors.

*Quantitative Evaluation:* We employed some of the widely used metrics in image colorization problem, namely the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), along with Weighted Path Histogram to quantitatively evaluate the performance of our models. Table 1 presents a summary of the scores we achieved on those metrics.

TABLE I: Evaluation score

| **Metrics** | **Model 1** | **Model 2** | **Model 3** |
|---|---|---|---|
| **PSNR** | 32.38 | 32.15 | 32.16 |
| **SSIM** | 0.96 | 0.96 | 0.96 |
| **Weighted Patch Histogram** | 0.73 | 0.75 | 0.77 |

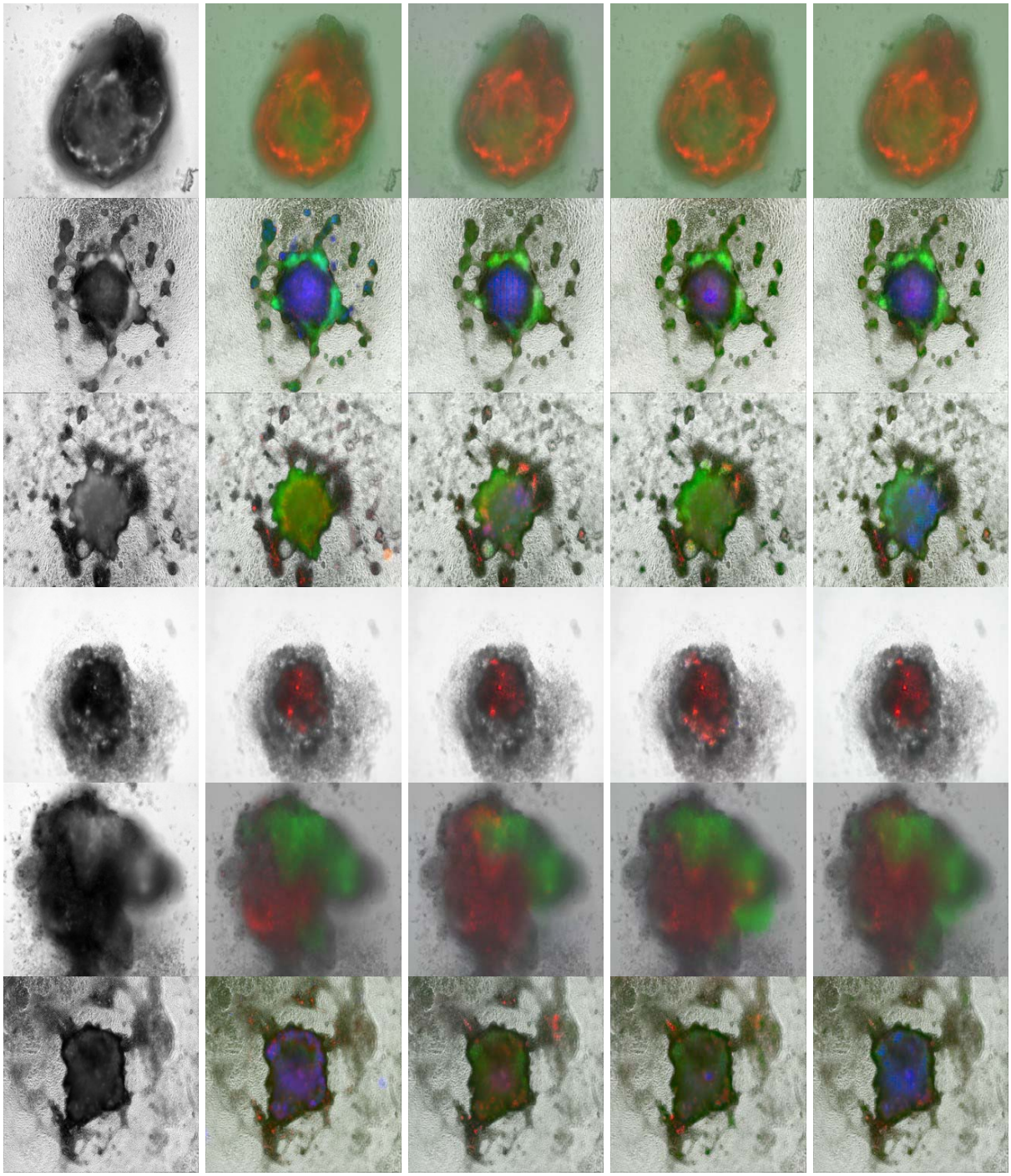
The range of Peak Signal-to-Noise Ratio (PSNR) is [0, ∞], where 0 represents no similarity between images and infinity is for exactly the same images, but for a comparison of lossy images, the PSNR score typically ranges between 30 to 50 where the higher the score higher the similarity [17]. Values over 40 are

usually considered to be very good and anything below 20 is unacceptable [2]. The state-of-the-art techniques achieved a PSNR score of 29.52 on the COCO-stuff dataset [18], whereas our models achieved PSNR scores are over 32.

The Structural Similarity Index (SSIM) score ranges in (1, 1] [15] where -1 represents no similarity and 1 represents very high similarity. Therefore, higher the score higher the similarity. The state-of-the-art techniques had an SSIM score of 0.94 on the coco-stuff dataset [18], whereas our models achieved SSIM scores of 0.96.

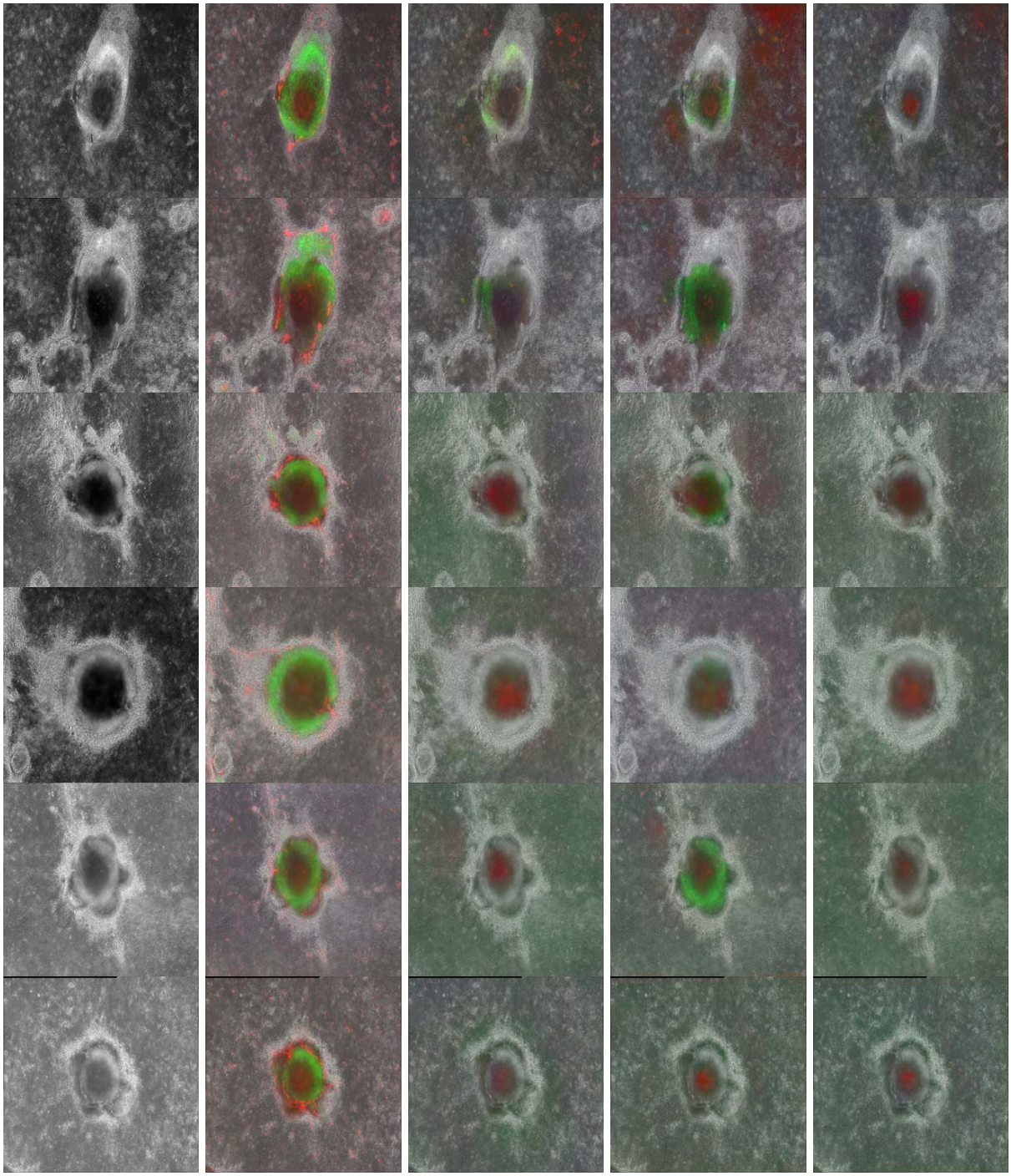
Weighted Patch Histogram ranges in [0,1] where 0 represents no similarity in the histograms of the images, therefore no similarity, and 1 represents full similarity in histograms resulting in a very high similarity of the images.

Beyond quantitative metrics, we present a qualitative assessment through visual comparisons. Figure 1 illustrates select input images from our biomedical dataset alongside the corresponding colorized outputs generated by our model. A similar set of images from the ImageNet dataset, colorized by the baseline model, is provided for reference.



(a) Input (b) Ground Truth (c) Model 1 Results (d) Model 2 Results (e) Model 3 Results

Fig. 5: (a) is the input, (b) is the ground truth, (c), (d), and (e) are the images generated by Model 1, Model 2, and Model 3 respectively



(a) Input (b) Ground Truth (c) Model 1 Results (d) Model 2 Results (e) Model 3 Results

Fig. 6: Colorization on new cell lines. (a) is the input, (b) is the ground truth, (c), (d), and (e) are the images generated by Model 1, Model 2, and Model 3 respectively

The images used in Figure 5, are for the organoids of the same cell lines but for the cells that were not included in the training dataset. We want to test the robustness of our models on the new cell lines that we did not have in the training dataset.

We applied our models to the cells obtained from new cell lines. Figure 6 shows the results generated by our models on the new cell lines. We can see that the generated colors are not as accurate as the ones we obtained with the predictions on the same cell lines in Figure 5. The results of evaluation metrics are shown in Table II informs us that the scores have dropped in terms of all the metrics. Visually only Model 2 is able to generate better results compared to the other 2 models. Therefore, we further fine-tuned Model 2 on the images from new cell lines to observe the color generation capability of our model.

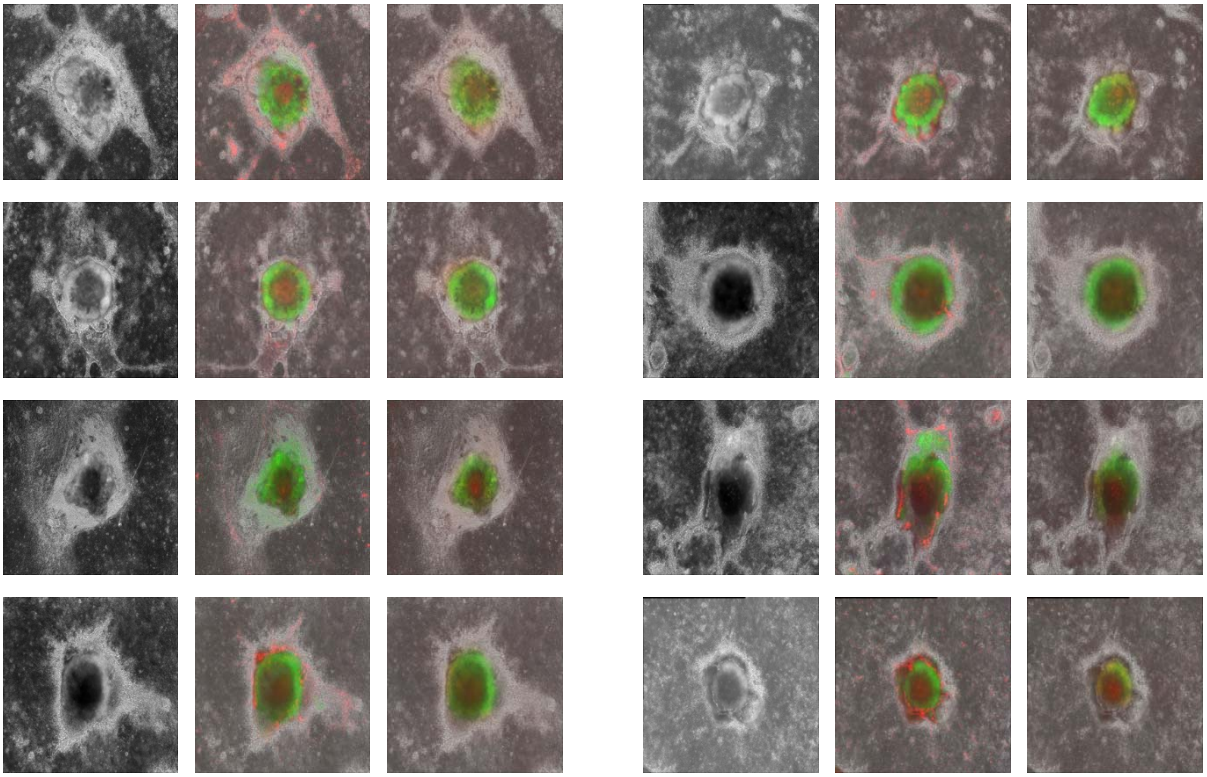
Figure 8 shows the results obtained on the new cell lines after fine-tuning and visually the color generation capability of our model has increased after fine-tuning. Also, from Table III we can see that the evaluation results have increased significantly after fine-tuning.

TABLE II: Evaluation score on new cell lines

| **Metrics** | **Model 1** | **Model 2** | **Model 3** |
|---|---|---|---|
| **PSNR** | 24.92 | 25.26 | 24.02 |
| **SSIM** | 0.95 | 0.92 | 0.92 |
| **Weighted Patch Histogram** | 0.49 | 0.52 | 0.44 |

TABLE III: Evaluation score on new cell lines after fine tuning

| **Metrics** | **Model 2** |
|---|---|
| **PSNR** | 29.82 |
| **SSIM** | 0.94 |
| **Weighted Patch Histogram** | 0.84 |



(a) Input (b) Ground Truth (c) Model 2 Results (a) Input (b) Ground Truth (c) Model 2 Results

Fig. 7: Fine tuning on new cell lines. (a) is the input, (b) is the ground truth, (c) is the images generated by Model 2

## CONCLUSION

In conclusion, our research has addressed the critical challenge of colorizing phase images of cardiac organoids, providing a solution that uses Conditional Generative Adversarial Networks (cGANs) and integrates the Convolutional Block Attention Module (CBAM). This framework has demonstrated its efficacy in capturing intricate color details within organoid cells, enhancing the interpretability and analysis of these structures in biomedical research.

Our GAN model, enriched by the CBAM module, outperforms the other two models, showcasing its adaptability and effectiveness in capturing intricate color details of cardiac organoids. Notably, for optimal results on new cell lines, finetuning the model is advisable, ensuring the

generation of accurate and faithful color representations.

Our evaluation, including the introduction of the Weighted Patch Histogram as a new metric, highlights the efficacy of our approach in preserving biological details compared to traditional metrics like PSNR and SSIM. The images generated with evaluation scores of PSNR over 30, SSIM over 0.92, and a Weighted Patch Histogram score over 0.75 are the most accurate and similar to the ground truth.

In future work, we will refine and expand our model to generalize on all cell lines and make the models generate colors directly from the Phase image rather than the grayscale image derived from the LAB colorspace. By continually refining our methods, we aim to contribute to the ongoing progress in organoid research and facilitate more accurate and meaningful analysis of cardiac organoids.

## REFERENCES

[1] 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.

[2] Bull, D. R. (2014). Digital picture formats and representations. Communicating pictures, 99-132.

[3] Fang, F., Wang, T., Zeng, T., & Zhang, G. (2019). A superpixel-based variational model for image colorization. IEEE Transactions on Visualization and Computer Graphics, 26(10), 2931-2943.

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[5] Gupta, R. K., Chia, A. Y. S., Rajan, D., Ng, E. S., & Zhiyong, H. (2012, October). Image colorization using similar images. In Proceedings of the 20th ACM international conference on Multimedia (pp. 369-378).

[6] He, M., Chen, D., Liao, J., Sander, P. V., & Yuan, L. (2018). Deep exemplar-based colorization. ACM Transactions on Graphics (TOG), 37(4), 1-16.

[7] Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2016). 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 (ToG), 35(4), 1-11.

[8] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

[9] Ji, G., Wang, Z., Zhou, L., Xia, Y., Zhong, S., & Gong, S. (2020). SAR image colorization using multidomain cycle-consistency generative adversarial network. IEEE Geoscience and Remote Sensing Letters, 18(2), 296-300.

[10] Kiani, L., Saeed, M., & Nezamabadi-pour, H. (2020, February). Image colorization using generative adversarial networks and transfer learning. In 2020 International Conference on Machine Vision and Image Processing (MVIP) (pp. 1-6). IEEE.

[11] Kuzovkin, D., Chamaret, C., & Pouli, T. (2015). Descriptor-based image colorization and

regularization. In Computational Color Imaging: 5th International Workshop, CCIW 2015, Saint Etienne, France, March 24-26, 2015, Proceedings 5 (pp. 59-68). Springer International Publishing.

[12] Larsson, G., Maire, M., & Shakhnarovich, G. (2016). Learning representations for automatic colorization. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14 (pp. 577-593). Springer International Publishing.

[13] Mouzon, T., Pierre, F., & Berger, M. O. (2019). Joint cnn and variational model for fully-automatic image colorization. In Scale Space and Variational Methods in Computer Vision: 7th International Conference, SSVM 2019, Hofgeismar, Germany, June 30–July 4, 2019, Proceedings 7 (pp. 535-546). Springer International Publishing.

[14] Nazeri, K., Ng, E., & Ebrahimi, M. (2018). Image colorization using generative adversarial networks. In Articulated Motion and Deformable Objects: 10th International Conference, AMDO 2018, Palma de Mallorca, Spain, July 12-13, 2018, Proceedings 10 (pp. 85-94). Springer International Publishing.

[15] Nilsson, J., & Akenine-Möller, T. (2020). Understanding ssim. arXiv preprint arXiv:2006.13846.

[16] Oza, U., Pipara, A., Mandal, S., & Kumar, P. (2022, July). Automatic Image Colorization using Ensemble of Deep Convolutional Neural Networks. In 2022 IEEE Region 10 Symposium (TENSYMP) (pp. 1-6). IEEE.

[17] Sara, U., Akter, M., & Uddin, M. S. (2019). Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. Journal of Computer and Communications, 7(3),

8-18.

[18] Su, J. W., Chu, H. K., & Huang, J. B. (2020). Instance-aware image colorization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7968-7977).

[19] Suárez, P. L., Sappa, A. D., & Vintimilla, B. X. (2017). Infrared image colorization based on a triplet dcgan architecture. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 18-23).

[20] Treneska, S., Zdravevski, E., Pires, I. M., Lameski, P., & Gievska, S. (2022). Gan-based image colorization for self-supervised visual feature learning. Sensors, 22(4), 1599.

[21] 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.

[22] Wang, Y., Xia, M., Qi, L., Shao, J., & Qiao, Y. (2022, October). PalGAN: Image colorization with palette generative adversarial networks. In European Conference on Computer Vision (pp. 271-288). Cham: Springer Nature Switzerland.

[23] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

[24] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

[25] Žeger, I., Grgic, S., Vuković, J., & Šišul, G. (2021). Grayscale image colorization

methods: Overview and evaluation. IEEE Access, 9, 113326-113346.

[26] Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14 (pp. 649-666). Springer International Publishing.

[27] Zhao, Y., Wang, L., Jin, W., & Shi, S. (2007, May). Colorizing biomedical images based on color transfer. In 2007 IEEE/ICME International Conference on Complex Medical Engineering (pp. 820-823). IEEE.