

A Comparative Survey on Grayscale Image Colorization using GAN Architecture

Dr. R. Mynavathi¹, Dr. P. Rajendran² and Karunambika S P³, Abishek S⁴, Jagadesh S K⁵ and Mafas Ahamed K⁶

^{1,3-6} Velalar College of Engineering and Technology/Information Technology, Erode, India

Email: rpgmyna@gmail.com, spkarunambika@gmail.com

² Velalar College of Engineering and Technology/General Engineering, Erode, India

Email: pnavya@gmail.com

Abstract—Grayscale image colorization is the process for estimating color values for Grayscale image and video frames to gain more insights of the target images and improve their quality aesthetically. Many colorization techniques have been broadly employed colorizations of old and historical photo, satellite images, scientific image illustrations, medical image processing to improve visual features and appeal. Over a decade, many approaches have been introduced for addressing various colorization issues. Existing methods widely falls into three categories namely scribble-based colorization, transfer-based colorization and Self-supervised colorization. This paper analyses various recent prominent works in the above categories. We analyze the reason for GANs being an effective model for image colorization and finally a recent method proposed for using image colorization and a proxy task using GAN is being reviewed to gain insights from the colorization and transferring the knowledge acquired in downstream tasks like segmentation and classification without manual annotation.

Index Terms— Image colourization, Automatic Gray scale image colourization, GANs, Conditional GAN, MSME, Peak Signal-to Noise Ratio, SSIM, LPIPS.

I. INTRODUCTION

Grayscale images lack detail and clarity compared to color images, hence in most of the applications where only black and white images are the output are frequently converted into a color images first[1,2]. Automatic conversion of gray scale images to plausible high-quality images are always interesting research area like historical image /videos restoration, image compression, medical images processing [3]. There are various reasonable color options to even a single object for example the balloon may be yellow, blue or even pink etc., this task can be considered multi modal colorization approach and it is a challenging problem and awaiting intriguing research. Traditional methods of colorization rely on user intervention like User-guided methods [4,5] and transfer-based methods[6]. Gray scale colorization methods for colorizing grayscale images includes user intervention to provide input or guidance, such as scribbles or reference images, to guide the colorization process. Self-supervised grayscale colorization refers to the process of automatically generating color images from grayscale images without using any additional supervision. These processes are based on the idea of leveraging the inherent structure of the data to learn meaningful representations in a self-guided manner without manual intervention.

Among several approaches to self-supervised methods for complex colourization task, GANs are considered a

prominent choice [3,11,12] because GANs can learn complex, high-dimensional distributions of data without the need for explicit labels[13]. This is particularly useful for image colorization, as it is a difficult task to solve using traditional supervised learning methods due to the lack of large-scale labelled data. The recent works introduced that knowledge acquired in image colourization using GAN[11] and other self-supervised models [14,15] can be transferred to further downstream tasks which helps to enhance the performance.

II. LITERATURE STUDIES

Each pixel's colour value on a grayscale image lies in the range of black and white. This section presents the various techniques for Gray scale image colourization. The grayscale colourization networks have been classified as User- Guided, Semi-Automatic and Fully-Automatic.

A. Scribble-based Colorization

Scribble-based colorization [5, 16] is a type of user-guided grayscale image colorization method. Here the hints are provided as input by the user in the form of color scribble on the image. Then the coloring optimization algorithm spreads the scribbled colors. The algorithm then uses these scribbles as a guide to fill in the remaining colors under the presumption that neighboring pixels with the equal pixel values would have similar colors.

The first work in scribble-based method proposed back in 2004 involved lot of human intervention and went through a larger improvisation over the years. The implementation of prior user-guided colorization works focused on user strokes which are propagated using low level metrics. Those system required user to provide scribble for an input image in an accurate manner which demanded high level knowledge from the users. In recent times, researches have been done to reduce this tedious process using deep learning networks.

Zhang et al. [5] suggested a Convolution Neural Networks (CNN) based model to simplify scribbling process by reducing manual intervention and produce better colorization outcomes. CNNs have been used in grayscale colorization, particularly as a supervised learning approach. This method combined the use of Deep Learning approach with user-guidance. This network propagates user edits instead of utilizing rules that are manually created by combining low-level cues with high-level semantic knowledge that has been learnt from massive amounts of data. The trained data includes millions of images along with simulated user inputs. This system also recommends likely colors to guide users towards efficient input color selection. Here the colorization task is finished in single feed-forward pass, which enables real-time use.

The two forms of user interaction enabled colorization network are trained namely Local user hint and Global hints. As shown in Fig1, the local hints analyses user input and forecasts the distribution of colors. Global statistics are included into the basic structure by the global hints network.

As the network allows the user to select the color for the grayscale images, reveals that if there is no adequate user hints this network can produce undesired effect on images being colorized and it may fail to entirely propagate a point intended by user. i.e., lesser the color points marked by user, the greater the chance of color bleeding.

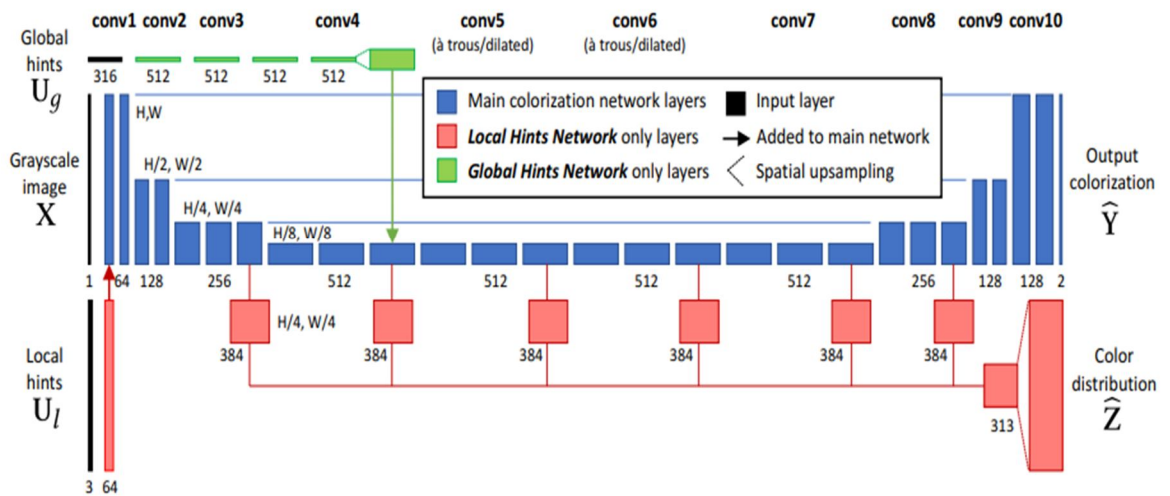


Figure 1. Network Architecture in User Guided Scribble Based Model

B. Transfer-based Colourization

To reduce manual intervention in the image reconstruction processes several methods like Transfer-based methods [6, 18]. have been introduced which required lesser manual effort from the users for image colorization. This typically involve transferring color information from a reference or source image to a grayscale target image. The main idea is to utilize color information in the reference image for guiding the colorization of target image. These techniques frequently rely on having access to the reference photos so that the colors may be translated to the desired grayscale input.

Hameed et al. [6] presented a simple and effective approach for black and white colorization. Associating the concept of image layer at different spatial frequencies have different physical and perceptual property with traditional color transfer-method proposed earlier by Welsh et al make the colorization process simple This content-based colorization method enhances the Welsh et al [17] method and provide more effective than the Welsh et al without any manual intervention for swatch pairs selection.

The performance of these methods will be affected by the similarity between reference and input images. Transfer-based methods can be effective only when the grayscale image (target image) has an identical form to reference images that are being used. However, they may not be as effective when the structure of grayscale image is greatly differed from the reference image.

C. Full-automatic Colourization

By adding semantics to the image, human beings can correlate grayscale images with colors, but it is challenging for them to infuse in hues. Manual carried out colorization tasks demands a lot of effort from the user. To reduce manually intervention self-supervised methods came into picture with numerous researches in the backdrop. They require less manual intervention and are better suited for complex image datasets.

Jiancheng An et al. [7] used the VGG-16 Convolution Neural network with images which can learn colors automatically without a user intervention. Here the input is fed into gray scale layer and then system predicts the layers in color space. Channel L (i.e., grayscale input) is used to obtain final color image. The mechanism for converting grayscale images in this network can be categorized into four steps:

- Firstly, the image in RGB is converted to color Lab color space.
- Then the L channel from Lab space is taken as input to the network which cause it to predict the ab (code a is for green and red and code b is for blue and yellow) channels.
- Thirdly, expected ab channels are combined with channel L
- Finally, the retrieved Lab image is transformed using conventional color space transforms into RGB.

Rebalancing after a classification loss produces more dynamic outcomes. The color balance can be used to enhance colors and can alter the entire color palette in an image. De-saturated values in photographs with backdrops like clouds, walls, roads, and dirt will cause a stronger bias of the loss function towards unsaturated values because ab values dispersed in the strongly influential way. This model corrects the problem of class imbalance by allocating a weight to individual pixel depend on rarity of its color. Re-sampling the color space using a multinomial loss function and color rebalancing is used to perform the reweighting.

However, the CNNs lack the ability of Data Generation and there is overreliance on training data which means CNNs demands a larger quantity of labeled data for training in order to learn to colorize grayscale images accurately. Also, there will be incomplete or biased training which will affect the colorization results. The lack of interpretability in CNNs which are often described as "black boxes" will affect the network's decisions.

Su et al. [8] gave a learning-based method to achieve object - based colorization. Instance aware colorization method is used get a clear separation of objects so that colorization performance can be improved. Here the network learns to colorize at objects (instances) level which is easier task on comparing to methods that makes the network to learn to colorize the image entirely and the instance-level colorization avoid color confusion with background and complex object.

This network consists of two branches of Colorization networks. One branch is for colorization of Instances and the other branch is for full image colorization. Main colorization network was adopted from the work of Zhang et al [5]. The two branches are enough to predict the color of instance image (Y_i) and full image (Y) the clueless blending results is visible artifacts because of overlapping pixels inconsistency.

The drawbacks of this network arises when the objects (instance) are not identified, the model returns backs to the usual form of network for image colorization. Then the output produced will have visible artifacts like object boundaries bleeding or color may appear washed out.

Convolution neural networks (CNNs) based methods have been used in gray scale image colorization, particularly as a supervised learning approach for a long time. However, CNNs have some limitations in

grayscale image colorization, such as difficulty in capturing long-range dependencies, sensitivity to small variations in input, and the tendency to produce over-smoothed outputs. To address these issues, other methods such as GANs [19, 20, 21] have been proposed for gray scale image colorization.

The GAN architecture, CNN architecture and other deep learning colorization methods have their own strength and weaknesses based on certain specification but GANs are generally considered better than CNNs in grayscale image colorization because they can generate more visually appealing and realistic colorizations. GANs are generative models that learn to generate new data from a given distribution, while CNNs are discriminative models that learn to classify input data into pre-defined categories.

D. Full-automatic Colourization using GAN

The GAN architecture used in grayscale image colorization involves generator and discriminator network. The generator network accepts in grayscale image and produces colorized output. The discriminator network takes in both the grayscale input and the colorized output image and tries to distinguish between them. Fig.2 depicts the general GAN based Architecture.

The generator network typically has encoder-decoder architecture with the skip connections. The encoder down-samples the input image while extracting features and the decoder up-samples the feature maps to produce output image. The skip connections help preserve high-frequency information from the input image.

The discriminator network is usually a Convolution Neural Network that takes in the concatenated grayscale input image and colorized output image and produces a binary classification indicating whether the output is real or fake.

During training, the generator is trained for producing colorized images that the discriminator cannot differentiate from ground truth images, while discriminator is trained for correctly classifying the ground truth and generated images. This adversarial training process leads to generator network from producing more realistic colorizations over time which make it hard for the discriminator to generated images.

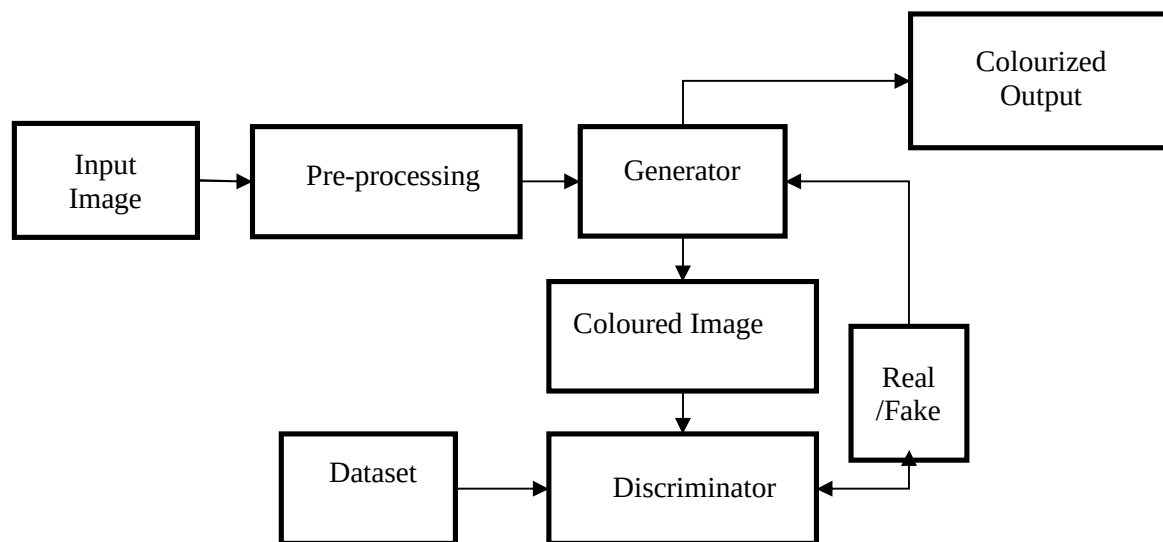


Figure 2 General GAN Architecture

Sankar R et al. [9] attempted to generalize the procedure of Colorization using Conditional Convolutional GAN by combining “Perceptual Loss” function. For the training of the network a dataset called CIFAR-100 is used. Here the base line system-built model of CNN using perceptual loss function is used. Then it is extended to the Adversarial Network. Colorization task is handled using LAB color space where one channel is dedicated for brightness portrayal and the other two channels is coded with color information to prevent the RGB intensities perturbations.

The Baseline network is CNN model which consists of convolutional layers for down sampling and up sampling. U-Net Architecture used to provide prominent edge locations. Architecture of generator and discriminator is modeled using Deep convolutional GAN(DCGAN) which is structured in CNN.

In generator Conditional GAN used to conduct experiments and dropout is employed for attaching noise for layers. Down sampling architecture is considered for discriminator with some necessary modification. The GAN

with perceptual loss obtained MAE of 8.312 which is less than both U-net and GANs. As a result of the comparative experimentation GAN combined with Perceptual Loss performs better than U-net and GANs.

Antic J et al. [10] introduced an open-source model based on deep learning approach. It uses the GAN architecture for colorizing the images. The new type of GAN model called NO-GAN is introduced. The main aim of the model is to obtain the advantages of GAN based training in lesser time with doing direct GAN training which requires a large amount of time. Here initially generator and the critic(discriminator) trained separately then finally both generator and discriminator are trained together in GAN setting. Deoldify trained 3 types(Artistic, Stable and Video) of model for images and video colorization. All of the model produced best result but there are discoloration problem arises in some target images and produce less colorful results. As the model learn from whole image colorization the result of deoldify suffer from context confusion in some target images

GANs achieve better results in colorization because they learn to generate color from context, while CNNs may have limitations in this regard. Additionally, GANs can learn to generate diverse colorizations for a single grayscale image, while CNNs may generate only one fixed output for a given input. When GAN are combined with perceptual loss it produce effective result in lesser amount of training. The GAN also enhance the performance of downstream task as well when used for image colorization as a proxy task Overall, GANs offer more flexibility and creativity in grayscale image colorization.

E. Fully-Automatic Grayscale colourization as Proxy Task

Grayscale image colorization is considered as a proxy task [22,11] because it helps to achieve a larger goal of generating realistic color images. By training a model on grayscale images and predicting color values, the model can learn to understand the underlying structure of images and generate color images with more realistic textures, patterns, and shapes. This approach can be more efficient and effective than directly training a model to generate color images from scratch, especially when working with limited data or complex image datasets. Additionally, other computer vision tasks including object recognition, image segmentation, and classification could also use grayscale image colorization as a pre-processing step.

Treneska S et al. [11] focused on image colorization with GAN due to their capability of producing the most accurate colorization results. This model is the first paper that train GAN for colorization as proxy task. It shows image colorization with cGANs can boost and support the performs of the other downstream task without manual annotation. This network proved that GANs can be used for automatic feature learning through image colorization. Image colorization task can be passed to multilabel image classification and segmentation.

By passing the learned weights to the other computer vision task like significant increase in the performance. In this paper the multilabel classification task started with combining the algorithm with the learned weights shown 5% better performance than training the classification algorithm alone. It has also 2.5% performance enhancement in the segmentation task that the baseline model. Fig 3 shows the flow of passing colorization weights to the multilabel image classification of PASCAL dataset.

In this work Image colorization is considered the target task. Conditional GAN used in based on Pix2Pix architecture. Here the U-Net Architecture in adopted for Generator which is designed for segmentation and discriminator is PatchGAN which gives the critic on the images that are real or fake.

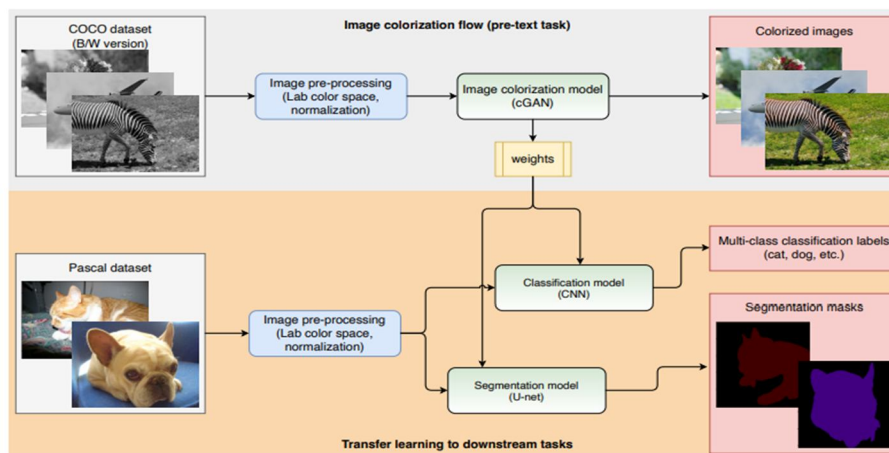


Figure 3 Dataflow of using image colorization as Proxy Task

III. COMPARISON OF RESULTS

TABLE II. COMPARATIVE STUDY OF RESEARCH ON GRAYSCALE IMAGE COLOURIZATION

Author	Title	Findings And Evaluation Metrics	Gap Identified
Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S. Lin, Tianhe Yu, Alexei A. Efros [2017]	Rea-Time User-Guided Image Colorization with Learned Deep Priors	- By minimising manual involvement, an instantaneous, manually controlled image colorization with a CNN can streamline the scribbling process and improve colorization results is proposed - Evaluation Metrics : PSNR and SSIM	Some cases reveal that in the missing of nearby user inputs could result an undesired effect on images being colourized and it may fail to entirely propagate a point intended by user.
Hameed, Sarab & Attea, Bara'a & Abbood, Amenah	Content-based Grayscale Image Colorization.	- This paper presents a method for associating the concept that image layer of different frequencies has different perceptual and physical properties with Welsh et al[16] for producing better result than traditional method - Evaluation Metrics : MSE	The performance of these methods will be affected by the similarity between reference and input images.
Jiancheng An, Koffi Gagnon Kpeyiton, Qingnan Shi [2020]	Grayscale images colorization with convolutional neural networks.	- It introduces an unsupervised learning-based system using classification with loss of cross-entropy VGG-16 CNN model. The colour rebalancing is done to colourized images that are more vibrant and have more plausible colours - Evaluation Metrics : RMSE	The CNNs lack the ability of Data Generation there is overreliance on training data. The inability of CNNs, which are frequently referred to as "black boxes," to be interpreted. This can make it difficult to fix or improve the model.
Su, J.W., Chu, H.K., Huang, J.B [2020]	Instance-aware image colorization	- A learning-based method using deep learning framework in order to achieve instance aware colourization is used. - Fusion module is newly introduced to obtain better mapping - Evaluation Metrics : PSNE,SSIM,LPIPS	The drawbacks of this network arises when the objects(instance) are not identified, the model returns back to the usual image colourization process. Then the output produced will have visible artifacts like object boundaries bleeding or colour may appear washed out.
R. Sankar, A. Nair, P. Abhinav, S. K. P. Mothukuri and S. G. Koolagudi [2020]	Image Colourization using GANs and Perceptual Loss	- With the addition of "Perceptual Loss," an effort was taken to generalize the colorization process using a conditional Deep Convolutional Generative Adversarial Network (DCGAN). - Evaluation Metrics : PSNR and SSIM	There is need for efficient quantitative metrics for evaluation
Jason Antic [2020]	A deep learning-based project for colorizing and restoring old images	- The NO-GAN model, a new form of GAN model, is presented. The model's primary goal is to reap the rewards of GAN training with the least amount of time spent performing direct GAN training. - Evaluation Metrics : PSNR and SSIM	There are discoloration problem arises in some target images and produce less colourful results. As the model learn from whole image colorization the result of deoldify suffer from context confusion in some target images
Treneska S, Zdravovski E, Pires I. M, Lameski P, Gievska S [2022]	GAN-Based Image Colorization For Self-Supervised Visual Feature Learning	- Image colourization is introduced as proxy task for understanding visual features. Conditional GAN is used to achieve the purpose of feature learning - Evaluation Metrics : PSNR and SSIM	Using CGANs require intensive amount of training using large scale dataset can increase the image colourization quality

IV. FINDINGS FROM THE RESULTS

From the above survey it is evident that using GANs like Conditional GAN along with adversarial loss functions and perceptual loss for image colourization can produce high quality and plausible result therefore they have better accuracy than only GAN and image colourization can be proxy task and knowledge from task can be transferred down to downstream tasks. There is 5% increase in downstream task like classification model when learning weights from GAN based algorithms used than the baseline models

The generator in GAN architectures is trained for producing colorized images that are same as true colour images, while discriminator is trained to differentiate the generated colorized images and the ground truth images. This adversarial process of training encourages the generator network to produce more realistic and accurate colorized images over time. Fig 4 shows that GAN like cGAN and GAN combined in perceptual loss perform better than CNN in case of Grayscale image colourization. In summary, both GANs and CNNs have strengths and weaknesses, and the choice of which model to use depends on the specific task at hand. But for the case of grayscale image colorization, GANs are often preferred over CNNs as they are able to generate more realistic and diverse colorized outputs.

V. CONCLUSION

In this paper, we explored colourization of Grayscale images using various methods from traditional methods to more recent using various deep learning model. It is shown that result obtained using GAN for image colourization are of high accuracy and object level colourization can produce more accurate results. GAN shows promising results in grayscale image colourization because it is able to learn complex nonlinear relationships between the grayscale input image and its corresponding colour image. GANs are able to generate visually appealing and realistic colorized images by training on large datasets of grayscale and colour images, and learning the distribution of the colour space. Using GAN based Image colourization as proxy task then the transferring learning from image colourization to downstream tasks can boost the performance of both downstream task and colourization.

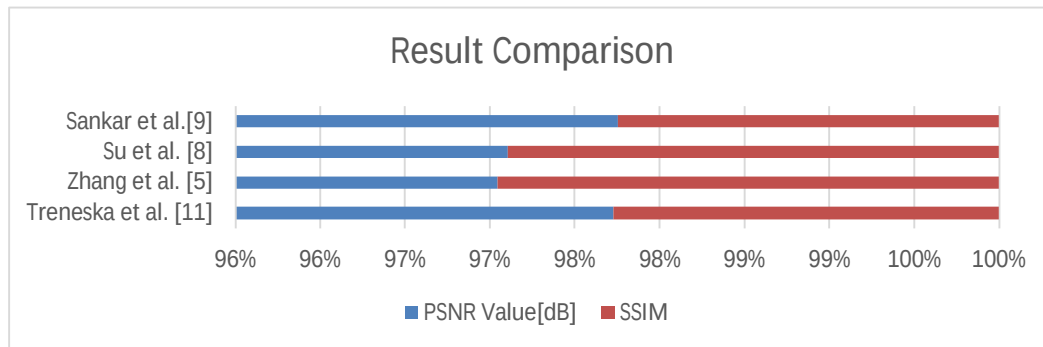


Figure 4 Performance Efficiency Comparison

REFERENCES

- [1] I. Žeger and S. Grgić, "An Overview of Grayscale Image Colorization Methods," 2020 International Symposium ELMAR, Zadar, Croatia, 2020, pp. 109-112, doi: 10.1109/ELMAR49956.2020.9219019.
- [2] Anwar, Saeed & Tahir, Muhammad & Li, Chongyi & Mian, Ajmal & Khan, Fahad & Muzaffar, Abdul. (2020). Image Colorization: A Survey and Dataset.
- [3] Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. *Med ImageAnal* 58:101552
- [4] Zhang, R.; Isola, P.; Efros, A.A. Colorful image colorization. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 649-666.
- [5] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros. Real-time user-guided image colorization with learned deep priors. *arXiv preprint arXiv:1705.02999*, 2017
- [6] Hameed, Sarab & Attea, Bara'a & Abbood, Amenah. (2019). Content-based Grayscale Image Colorization.
- [7] An, J., Kpeyton, K.G. & Shi, Q. Grayscale images colorization with convolutional neural networks. *Soft Comput* 24, 4751-4758 (2020). <https://doi.org/10.1007/s00500-020-04711-3>
- [8] Su, J.W., Chu, H.K., Huang, J.B.: Instance-aware image colorization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7968-7977 (2020)
- [9] R. Sankar, A. Nair, P. Abhinav, S. K. P. Mothukuri and S. G. Koolagudi, "Image Colorization Using GANs and Perceptual Loss," 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, 2020, pp. 1-4, doi:10.1109/AISP48273.2020.9073284
- [10] Jason Antic jantic/deoldify: A deep learning based project for colorizing and restoring old images (and video!). <https://github.com/jantic/Deoldify>. Online (2020).
- [11] Treneska S, Zdravevski E, Pires I. M , Lameski P, Gievska S, "GAN-Based Image Colorization for Self-Supervised Visual Feature Learning", *Sensors* 2022,22,1599. Feb 2022.

- [12] Nazeri, K.; Ng, E.; Ebrahimi, M. Image colorization using generative adversarial networks. In Proceedings of the International Conference on Articulated Motion and Deformable Objects, Palma de Mallorca, Spain, 12–13 July 2018; pp. 85–94.
- [13] Cao, Y.; Zhou, Z.; Zhang, W.; Yu, Y. Unsupervised diverse colorization via generative adversarial networks. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Skopje, North Macedonia, 18–22 September 2017; pp. 151–166.
- [14] Jing, L.; Tian, Y. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 43, 4037–4058.
- [15] Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
- [16] Patsorn Sangkloy, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. 2017. Scribbler: Controlling Deep Image Synthesis with Sketch and Color. *CVPR* (2017).
- [17] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. Transferring color to greyscale images. *ACM TOG (Proc. SIGGRAPH)*, 21(3):277–280, 2002.
- [18] Li B, Lai Y, John M, Rosin PL (Sep. 2019) Automatic example-based image colorization using location aware cross-scale matching. *IEEE Trans Image Process* 28(9):4606–4619. <https://doi.org/10.1109/TIP.2019.2912291>
- [19] Han Zhang, Ian Goodfellow, Dimitris Metaxas, Augustus Odena, “Self-Attention Generative Adversarial Networks”, Proceedings of the 36th International Conference on Machine Learning, PMLR 97:7354-7363, 2019
- [20] Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
- [21] Vitoria, P.; Raad, L.; Ballester, C. Chromagan: Adversarial picture colorization with semantic class distribution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 2445–2454.
- [22] Kiani, L.; Saeed, M.; Nezamabadi-pour, H. Image Colorization Using Generative Adversarial Networks and Transfer Learning. In Proceedings of the 2020 International Conference on Machine Vision and Image Processing (MVIP), Qom, Iran, 18–20 February 2020; pp. 1–6.