ISSN 2278-3091
Volume 10, No.4, July - August 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse401042021.pdf
https://doi.org/10.30534/ijatcse/2021/401042021

Image Colorization Progress: A Review of Deep Learning Techniques for Automation of Colorization

Harish Dalal, Anagha Dangle, RadhikaMundada, JeenishaShrungare, Sonal Gore
Department of Computer Engineering, PimpriChinchwad College of Engineering, Pune

ABSTRACT

Image colorization is the process of taking an input gray-scale (black and white) image and then producing an output colorized image that represents the semantic color tones of the input. Since the past few years, the process of automatic image colorization has been of significant interest and a lot of progress has been made in the field by various researchers. Image colorization finds its application in many domains including medical imaging, restoration of historical documents, etc. There have been different approaches to solve this problem using Convolutional Neural Networks as well as Generative Adversarial Networks. These colorization networks are not only based on different architectures but also are tested on varied data sets. This paper aims to cover some of these proposed approaches through different techniques. The results between the generative models and traditional deep neural networks are compared along with presenting the current limitations in those. The paper proposes a summarized view of past and current advances in the field of image colorization contributed by different authors and researchers.

Key words : Image colorization, Generative Adversarial Network, Convolutional Neural Network, Deep Learning.

1. INTRODUCTION

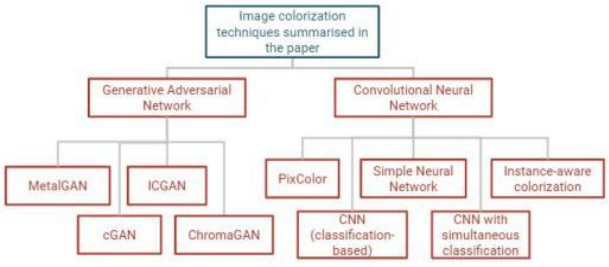

Figure 1: Various Image colorization techniques

Colorization is the process of processing an input gray-scale image and converting it into a colorized image that represents the colors and tones of the input. In the history of photography, if you had to see a black and white photograph and in color, pigment needed to be added with a paintbrush. Colorizing images that were taken hundreds of years ago allows us to see history from a new perspective, which adds a completely different layer of understanding the scene in the picture. Manual or traditional colorization, however, requires patience, is tedious and time-consuming. It is also prone to human error, and then the image can no longer be revoked. Recently though, digital methods, and now artificial intelligence-driven methods, have made colorization available to everyone. Digital colorization is a crucial task in areas like medical diagnosis, film industries, artist assistance, cultural heritage, and even crime prevention. The practice of colorization of black-and-white photos and videos has been on the increase during the last ten years, as image-editing software has become ever more prevalent and advanced. Colorizing archival images offers an opportunity to feel somehow more emotionally connected to a person or place and also go back in time to relive those moments. Tending to the need for automatic image colorization, areas under machine learning and image processing have come up with several algorithms and methods for the same. Though there are methods such as the ‘Scribble method’ which requires the user to manually mark color scribbles on the target image; still defies the aim of fully automated image colorization. Many existing models semi-automatic and automatic have succeeded in colorizing comics characters, historical photographs, sketches, and scenic images. Some of the machine learning networks namely CNN (Convolutional Neural Networks) (Ref. Fig 2) and GAN (Generative Adversarial Network) (Ref. Fig 5) are being briefly discussed in this paper(Ref. Fig 1). Both architectures object patterns to divide the target image into segments and add the required color to it accordingly. Considering a couple of different approaches for the same, this paper represents a literature survey, use cases, results, and conclusion. The research in this field has been branched into varied applications such as Infrared Image Colorization and Night vision image colorization. In addition to this, many researchers are focusing their interest on video colorization. Video Colorization can add insight features to many grayscale videos and can help in

preserving many historical events. It can be helpful to colorize the Gray-scale footage of CCTV and thus can help in precise observation.

2. LITERATURE SURVEY

A. Matí asRichart et al [1]

The paper [1] proposed a simple method centered on using back propagation to train a simple classifier over a training collection of color and grayscale images. The first step is the image preprocessing, which mainly consists of taking a color RGB image and converting it to the CIELUV [2] color space. Based on the gray levels of the pixels around it, the classifier predicts the color of a pixel. A local texture is captured in this tiny patch. The colors are reduced using Self-Organizing Maps [3] to keep the predictor's domain minimal. This reduction yields a limited collection of chroma values with enough variance to approximate all of the colors in the training set.

A Multi-Layer Perceptron Neural Network is used in the proposed method to predict the color of gray level pixels by classification. The paper suggests a vector quantization approach based on Self-Organizing Maps (as it is a fast and self-learning alternative [4]) to minimize the search space and organize colors in a limited number of clusters because the number of possible colors to learn is large. The neural network then divides gray-level pixels into color clusters.

To reduce the number of possible values to learn, the paper proposed to reduce the color range and for it, the technique of vector quantization is used. The Self Organizing Maps (SOM) [3] is used for vector quantization and for it the paper used sompy [5]. The size of the SOM is an important aspect in prediction. Lower the SOM's size higher the color reduction in the output image and Higher the SOM's size has a heavy impact on the training as it relates directly to the output space. As a result, this paper chose to use a 9-neuron SOM.

The dataset used in this paper is the "Open Country" image collection from the Label Me project [6]. Although the proposed method does not produce the same color results as more complex works, it is thought to be very effective due to its simplicity.

B. Iizuka et al. [7]

This paper [7] proposed a novel architecture shown in Fig. 2 that can jointly extract global and local features from an image and then fuse to perform the final colorization which is based on Convolutional Neural Networks. This method neither requires preprocessing nor post-processing, everything is learned in an end-to-end fashion. The model consists of four main components: a low-level features network, a mid-level features network, a global features network, and a colorization network. The Mean Square Error(MSE) criterion is used to train the network. As the model has a side product of the classification model and the training dataset is a large-scale dataset for the classification of N classes, thus it can be used to guide the training of the global image features.

The classification network is trained using the cross-entropy loss, jointly with the MSE for the colorization network. The model has trained on the Places scene dataset [8], consisting of 2,448,872 training images and 20,500 validation images. (compared with the state of the art [9]). The method's main drawback is that it is data-driven, which means it can only colorize images that have similar properties to those in the training collection.

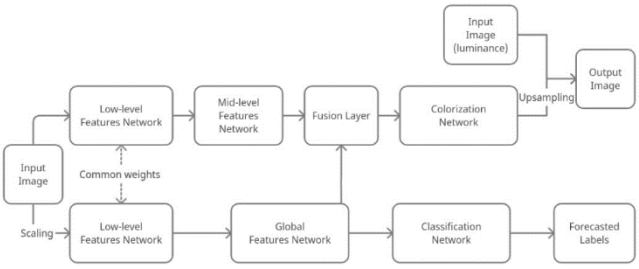

Figure 2: Iizuka et al. [7] proposed model architecture

C. JianchengAn* et al. [10]

This paper [10] used VGG-16 CNN [11] with 8 Convolutional Networks Layers each layer consisting of 2 or 3 Convolutional blocks followed by a ReLu and ending with a Batch Normalization layer [12]. The paper experiments the model used with a cross-entropy loss which is generally used with the classification-based approach.The input grayscale image is resized to 224 x 224. The multimodal cross-entropy loss function is used which is changed to include color rebalancing which allows the image to be more vivid and realistic in appearance. The problem of class imbalance in this classification work is solved by reweighting the loss of each pixel based on pixel color rarity. This reweighting can be affected by re-sampling the training space. The model was trained on ImageNet because of the variety of images that enabled the model to quickly learn different colors. Although the VGG-16 was used as a basic reference, ImageNet is used to train the model because the data set includes various images allowing our model to easily learn the different colors. The proposed image Colorization approach has been implemented in Caffe [13] and is capable of colorizing the grayscale images fully automatically.

The models were trained on 1,000,000 color images taken on the ImageNet and the result for the proposed model is 6.8069 per pixel RMSE (Root Mean Square Error).

D. Jheng-Wei Su et al. [14]

In this approach, the authors have proposed a novel approach by using an approach to detect variable instances in the given image. The architecture here extracts the feature from the instance branch and full images branch, then fuses

the both features together with the newly proposed fusion module and obtains a better feature map to predict the better results.

The architecture is experimented with by using various datasets such as ImageNet [15], COCO-Stuff [16], and Place205 [17].

The flow of the model in Fig 3 is as follows: Instance-aware processing provides a clear figure-ground separation and facilitates synthesizing and manipulating visual appearance. The system takes a grayscale image as input and predicts its two missing color channels in the CIEL*a*b* color space in an end-to-end fashion. The model leverages the pre-trained off-shelf object detection model Mask R-CNN [18]. After detection, the instance is cropped and scaled to 256 x 256 resolution. The instance image and the original image are passed for extracting the feature map by using the Real-Time User-Guided Colorization, main colorization network introduced in Zhang et al. [19] The feature maps of both consist of channels of 13 layers with number of channels 64, 128, 256, 512, 512, 512, 512, 256,256, 128, 128, 128 and 128. The feature maps of variable instances and full image fed to a network of 3 convolutional layers to get the weighted map of both by using ADAM Optimizer [20]. The feature map of variable instances is then fed to box bounding; the sole purpose of box bounding is to resize the instance feature according to the size of the full image and perform zero-padding on both. To fuse the features of both instance and full image softmax normalization is applied.

The model gives a Peak signal-to-noise ratio(PSNR) of 27.562 for the dataset of ImageNet, 28.522 for the dataset of COCOStuff validation split, 27.800 for the dataset of Places205 validation split.

The Input GrayScale Image is cropped according to the objects detected in Mask R-CNN[18] method. Then, the feature maps obtained are Full-Image Colorization and Instance Image Colorization by Real-Time User-Guided Colorization, the main colorization network introduced in Zhang et al. [19]. The Fusion model then fuses all the instances' feature maps in every layer with the extracted full-image feature map and thus obtains a full colorized image.

E. Sergio Guadarramaet. al [21].

This paper proposes to employ a conditional Pixel-CNN [22] to produce a low-resolution color image 28 x 28 from a given grayscale image 224 x 224 . Then, a refinement CNN is then trained to process the original grayscale and the low-resolution color image to produce a high-resolution colorized image. The model brings out multiple versions of input from low resolution to high resolution.

The approach can be formalized as

$$p(y\backslash - x) = \sum z\delta(y=f(x,z))p(z\backslash - x)$$

where x is the input grayscaleimage,y is the output color image. PixelCNN estimates p(z|x) is generated by PixelCNN and the refinement CNN estimatesy=f(x,z).

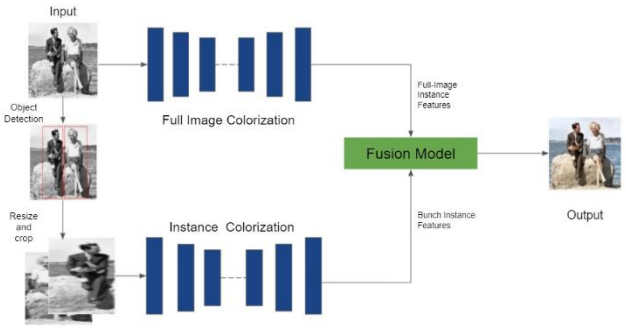

Figure 3: Diagram for Instance Image colorization

Technique. The In-put GrayScale Image is cropped according to the objects detected in Mask R-CNN[18] method. Then, the feature maps obtained are Full-Image Colorization and Instance Image Colorization by Real-Time User-Guided Col-orization, the main colorization network introduced in Zhang et al. [19]. The Fusion model then fuses all the instances' feature maps in every layer with the extracted full-image feature map and thus obtains a full colorized image.E

This architecture is based on [23] who used PixelCNNs to perform super resolution.The colorization of an individual pixel is determined by the color of the previous pixels. PixelCNN is composed of two convolutional layers and three ResNet blocks [24].The PixelCNN colorization network is composed of three masked convolutional layers: one in the beginning and second at the end of the network, whereas a Gated convolutional Block with ten layers is surrounded by the Gated convolutional layers.

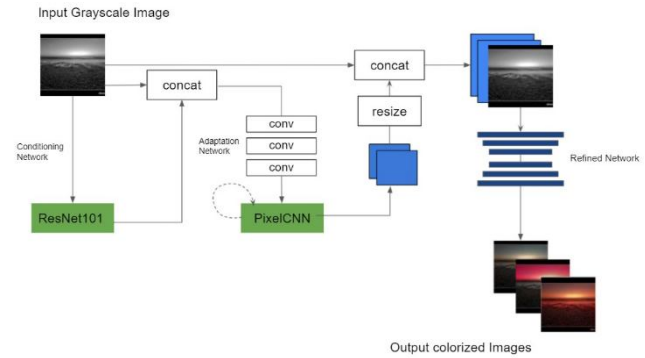

Figure 4: Model of Pixel Recursive Colorization Method.

The conditioning network is pre-trained.Then,the conditioning network and the adaptation network convert the input image brightness channel Y into a set of fea-tures providing the necessary conditioning signal to the PixelCNN.Then, PixelCNN predict low resolution images. The low resolution image is sub-sequently supplied to a refinement network, which is trained to produce a full colorization.

The PixelCNN model is trained by applying a maximum likelihood function with cross-entropy loss. During training, all the previous pixels are clamped to the ground truth values (an approach known as ”teacher forcing” [25]), and the network is trained to predict a single pixel at a time. The network ends with three convolutional layers, where the final layer outputs the colorized image. The Model is tested over ImageNet[15] test set which gives VTT(Visual Turing Test) Average score of 33.9 percent.

F. Nazeri et al. [26]

This paper by Nazeri et al. [26] provides an approach for the method of colorization using Generative Adversarial Networks proposed by Goodfellow et al. [27]. The network is trained on the datasets of CIFAR-10 and Place365 [28] and the results are later compared with those obtained using traditional convolutional neural networks (CNN). The architectures of generator and discriminator (two smaller networks of GAN) both follow a multi-layer perceptron model. The paper proposes a modified generator's cost function by maximizing the probability of the discriminator being inaccurate. As opposed to the traditional GAN, a variant called conditional GAN [29] is used so that the discriminator gets colored images from the generator as well as the original data of grayscale images as the condition. The L*a*b* color space is used to prevent the problem of sudden color and brightness that are experienced in RGB color space.

The paper suggests the use of a baseline model similar to that of U-Net architecture [30], where the discriminator has a series of convolutional layers with batch normalization. All the convolutional layers are followed by batch normalization [31] and leaky ReLU [32] activation with a slope of 0.2. After the last layer, the output is returned as a probability value of the input being real or fake. The network was trained using Adam optimization [20] and weight initialization as proposed by [33]. To make the training stable guidelines from [34, 35, 36, 37] are followed by the authors.

One of the drawbacks mentioned in the paper is that the GAN tends to colorize the objects in colors that are more frequent in the data set. However, the training result images from GAN showed significant visual improvement compared to those from baseline CNN.

G. Patricia Vitoria et. al. [38]

This paper [38] proposes an end-to-end adversarial learning approach with semantic class distribution learning called ChromaGAN to colorize a grayscale image. The generator of the ChromaGAN is divided into three stages in which the first stage has a structure similar to VGG-16 without the three last fully connected layers at top of the network. These layers are initialized with the pre-trained VGG-16 [39] weights which are not frozen during training. After this first stage, the sub-networks are split into two branches.

The first branch processes data using Convolutional layers followed by Batch normalization [12] and ReLU. The second track processes data using four modules of the form Conv-BatchNorm-ReLU, followed by three fully connected layers which output the class distributions. The third layer fuses both outputs from these branches. This then passes through six modules of the form Convolutional-ReLU with two up-sampling layers in between. The discriminator is based on the Markovian discriminator architecture(PatchGAN [40]). In this, the discriminator focuses on the local patches, and instead of giving a single output for the full image, it classifies the patches as real or fake.

The data set used for training consists of subsets taken from ImageNet [41] and are trained using ADAM optimizer for 5 epochs. This method outperforms the other state-of-art methods in terms of perceptual qualities in quantitative comparisons.

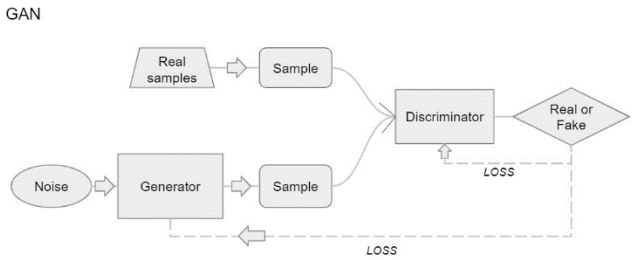

Figure 5: Generative Adversarial Networks Architecture

H. Paulina Hensmanet. al. [42]

This paper[42] involves the colorization of the popular Japanese graphic art, known as Manga [43] [44]. Colorizing using cGAN (Conditional General Adversarial Network) usually produces blurred results which can be rectified using color correction and segmentation, which in turn produces sharper, high-resolution images. The method described in this paper [45] uses a single colorized reference image as a training dataset; facilitating the ease of choosing a color for images involving the same character for example. The training data for cGAN colorization comprises a colorized image and its corresponding monochrome original. The input data is the monochrome target image to be colored. Datasets used in this particular paper are Monster Beat and Morevna [46], both of which are manga graphic comics.

The first step calls for screentone removal [47]. As screentone removal in high-resolution images is not perfect due to its elevated sharpness, it is preceded by the Gaussian blur method (with kernel radius of 1 or 2) to help with the separation of tone and edge lines. Using reference images similar to the target image will be more efficient. (which can be done by mapping images using edge lines for objects. Here, body parts.) For example, if the target image is a face,

the reference image being a similar face would be beneficial, for which cropping should be done if required. It is recommended to either train a single network with different crops of images or training multiple networks based upon several crops of the provided reference image.

The further action is segmentation, which is responsible for differentiating the segments or parts of the target image, each of which will differ in color, based on the edge lines. 'Trapped ball segmentation' [48] is a method used that delivers clearer and sharper segments, preventing leakage. (2 to 5 pixels in diameter) The main colorization occurs in 5 steps namely: Selecting color per segment, Increase saturation, color quantization, and Shading. Color selection is done by choosing the mean RGB palette color of all the corresponding pixels in the reference image and offers less complexity. Due to the colors seeming washed out because of the previous step, saturation is increased to strengthen the color of a segment and make it distinguishable from other segments (Values in the range 5-10 percent) are known to produce better results). color Quantization though not being a necessary step always, helps to reduce the number of different colors used in a small part of the segment. If the target image has screen tones, it easily qualifies to have shadings in its output image. After applying a Gaussian blur, each pixel of the appropriate area is then darkened to produce the shading effect. This being the last post-processing step, now the colored image is corrected with blurred edge lines and has much-minimized color spillage between segments. This method poses to be efficient because of its single image training resulting in increased training time. The post-processing techniques of color correction and segmentation reform the final image of blurring, color leakage, and other problems.

I. TomasoFontaniniet. al. [49]

Unlike most colorization algorithms that involve the system's training over large datasets, this paper[49] aims on providing just as good colorization results without the need for huge training datasets. Emphasizing the use of few-shot learning and one-shot learning, the network is then responsible to have a high generalization capacity because of the limited training datasets. The paper promotes the concept of 'learning to learn or meta-learning by adding a meta-layer on top of the traditional learning layers. Tasks are defined as clusters (natural grouping over data) of the initial dataset and are required to perform the same task over the same dataset, but differing at practical level functions. For example, a task would focus on the coloring of the 'A' type of objects, while another would focus on the coloring of the 'B' type of objects. Using Convolutional Neural Networks (CNN) the features are extracted for dividing tasks, and images clustered through K-means (A form of vector quantization to segregate n observations into k clusters in which each observation belongs to the cluster with the nearest mean).

Each cluster is then trained and designed to perform some designated tasks. The ambiguities of the basis of clusterization to meaningfully distribute tasks, ascertaining the efficiency of task specialization for coloring, the correct use of a meta-learning algorithm, and few-shot colorization for the colorization method to stand out among the traditional methods, persists. To tackle the above problems, a new architecture that combines meta-learning techniques and Conditional GANs known as 'MetalGAN' is proposed, and it specifies how the generator and discriminator parameters are updated. Clusterization to tackle image-to-image translation problems, and how good colorization results can be achieved even with a small dataset is also underlined.

The very first step of the algorithm involves the clusterization of the dataset used to distinguish the dataset into separate tasks. Using the activation 43 layer of Resnet50 (a variant of the Resnet model with 48 convolution layers, 1 max pool, and an average pool layer). By applying max pooling and L2 normalization on the features, we calculate the MAC descriptors, on which Principal Component Analysis (PCA) is applied to reduce feature dimensions and then apply K-means to divide the dataset into k clusters and tasks. Next, U-net is chosen to build the discriminator following DCGAN architecture [37], which has modules composed of convolutions, batch normalization, and ReLU layers. Concatenating the grayscale image xi input to the generator and the output ab channels, we obtain the final results. L1 loss[50] and adversarial loss are used to model low-frequencies and high frequencies of the output images, respectively. The activation function is formulated, where the weights are assigned to the different losses because we want the L1 loss to be more effective than adversarial loss during training. Once the task is chosen, the gradient of generator loss functions to perform the SGD (Stochastic Gradient Descent) step of optimization. On updating the inner loop parameters in the generation of the task, they are aligned with the Reptile rule. Therefore, the paper explains how the networks trained with only a few images can still produce good colored images with the use of the MetalGAN technique.

3. LIMITATIONS AND FUTURE WORK

We studied different Deep Neural Techniques for Image Colorization and summarized it in Table I and Table II. Our study revealed the following: Different colorization papers typically employ varied evaluation metrics which are not uniform, hence comparative study based on these metrics becomes a difficult task. Also, different models employ different color channels based on the techniques used. Image colorization techniques are typically evaluated on gray-scale images from various datasets available on the internet due to the absence of datasets built for this specific task. The available datasets were originally collected for tasks such as detection, classification, etc., and not specifically for image colorization and hence the quality of the images may not be sufficient for image colorization. Also, these datasets have the same object in different colors and hence it poses problems during the colorization task. Also unlike the object detection and classification works, image colorization is not

widely popular in the open-source community. There is a lack of resources be it videos or tutorials to use the models and hence the progress is hindered to some content.

Many models are implemented using limited datasets to obtain fine results. There may be room for development in fine coloring the background of a complex image. Images containing many instances are complex to colorize. Research can be improvised or shifted towards hybridization of different technologies and by experimenting with different feature extraction techniques.

Though the images are being colorized in proper shades of pigments, they sometimes lack the real sense and may contain a fewer number of colors as compared to the original. Research is going on in producing standardized datasets and metrics for evaluation. The future of image colorization promises a more realistic production of these gray-scale pictures. The research is also extending in the field of video colorization which finds its applications in surveillance, etc.

Table 1: Comparative study of deep learning techniques for image colorization

Authors	Objectives	Work details	Dataset	Acuracy
MatíasRichart, Jorge Visca, Javier Baliosin [1]	Image Colorization with Neural Networks	Neural Network approach, trained. Color range is reduced with the technique of vector quantization.	”Open Country” image collection from the Label Me project.	Less than 50% of the more complex approaches.
Satoshi Iizuka, Edgar Simo-Serra, Hiroshi Ishikawa [7]	Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification.	Novel architecture that extract global and local features from image which are fused for final colorization	Places scene dataset.	Naturalness (median) 1.Ground truth - 97.7% 2.Proposed - 92.6%
Jiancheng An, Kpeyiton Koffi Gagnon, Qingnan Shi, Hongwei Xie, Rui Cao [10]	Image Colorization with Convolutional Neural Networks	VGG-16 with a multiple convolutional blocks	ImageNet	Per-pixel RMSE - 6.8069 Survey (12 people) - 4.5 / 6
Jheng-Wei Su, Hung-Kuo Chu and Jia-Bin Huang [14]	Instance Aware Image Colorization	The instance detected and full image are colorized individually. Later both the feature maps are fused.	ImageNet, COCOStuff, Places205	PSNR 0.134, PSNR 0.125, PSNR 0.130
Sergio Guadarrama, Ryan Dahl, David Bieber, Mohammad Norouzi, Jonathon Shlens, Kevin Murphy [21]	PixColor : Pixel Recursive Colorization	Low resolution colored image is generated by conditional PixelCNN	ImageNet	Average VTT score: 33.9%
KamyarNazeri, Eric Ng, and MehranEbrahimi [26]	ICGAN	GAN following a multi layer perceptron model.	CIFAR-10 dataset ,Places365 dataset	1 MAE-5.1 (CIFAR-10) 2.MAE :7.5 (Places 365)
Patricia Vitoria, Lara Raad and Coloma Ballester [38]	ChromaGAN	GAN with semantic class distribution learning.	ImageNet	PSNR (dB) - 24.98
Paulina Hensman and Kiyoharu Aizawa [42]	cGAN colorization	Using colorized reference and corresponding monochrome image for training, followed by post processing techniques	pages of the manga Monster Beat ,Frames from project Morevna	Inception score Std:3.20 Mean: 0.83
TomasoFontanini,	Metal-GAN	Combination of Conditional	Mini-Imagenet	Inception

EleonoraIotti and Andrea Prati [49]	GAN with meta learning techniques.	score Std:9.16 Mean:1.12

5. CONCLUSION

Image colorization is a research problem with critical real-life applications. Deep learning approaches' exceptional success has resulted in rapid growth in deep convolutional techniques for image colorization. Based on current advancements, various methods are proposed with varied network structures, training methods, and learning paradigms, etc. In this paper, we have reviewed and compared these different architectures to provide a summary of the research in this field. While researching we have observed that image colorization performance has improved in recent years. However, this method's application to critical real-world scenarios is restricted due to inadequate metrics and network complexity. Recent trends in image colorization have shown the extensive use of GAN-based methods to deliver diverse colorization visually compared to CNN-based methods. There have been works to improve the models with higher complexity to provide better results.

REFERENCES

[1] *M. Richart, J. Baliosian, J. Visca, "Image Colorization with Neural Networks", IEEE Natal, Brazil, 2018.*J. U.

[2] ISO/CIE 11664-5:2016 - Colorimetry – Part 5: CIE 1976 L*u*v* color space and u', v' uniform chromaticity scale diagram, International Organization for Standardization, Geneva, Switzerland. Std., 2016.

[3] T. Kohonen, "Self-organized formation of topologically correct feature maps," Biological cybernetics, vol. 43, no. 1, pp. 59–69, 1982.

[4] J. Yoo and S.-Y. Oh, "A coloring method of gray-level image using neural network," in Proceedings of the 1997 International Conference on Neural Information Processing and Intelligent Information Systems, vol. 2, 1997, pp. 1203–1206.

[5] V. Moosavi, S. Packmann, and I. Vallés, "SOMPY: A Python Library for Self Organizing Map," 2016. [Online] Available: https://github.com/sevamoo/SOMPY

[6] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International Journal of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001. [Online]. Available: http://dx.doi.org/10.1023/A:1011139631724.

[7] S. Lizuka, E. Simo-Serra, H. Ishikawa, "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification" , ACM Transactions on Graphics, July 2016.

[8] Z HOU , B., L APEDRIZA , A., X IAO , J., T ORRALBA , A., AND O LIVA , A. 2014. Learning deep features for scene recognition using places database. In NIPS.

[9] C HENG , Z., Y ANG , Q., AND S HENG , B. 2015. Deep colorization. In Proceedings of ICCV 2015, 29–43.

[10] J. An, K. K. Gagnon, Q. Shi, H. Xie, R. Cao, "Image Colorization with Convolutional Neural Networks", IEEE Suzhou, China, 2020.

[11] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556, 2014.

[12] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. CoRR, abs/1502.03167, 2015.

[13] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to grayscale image," ACM Transactional on Graphics, vol. 21, no. 3, pp. 277-280, 2002.

[14] Su, Jheng-Wei, Hung-Kuo Chu, and Jia-Bin Huang. "Instance-aware image colorization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[15] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, San jeevSatheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, AdityaKhosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge.IJCV, 115(3):211–252, 2015.

[16] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In CVPR, 2018.

[17] Bolei Zhou, AgataLapedriza, Jianxiong Xiao, Antonio Tor-ralba, andAudeOliva. Learning deep features for scene recognition using places database. In NIPS, 2014.

[18] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross B.Girshick. Mask r-cnn. In ICCV, 2017.

[19] Richard Zhang, Jun-Yan Zhu, Phillip Isola, XinyangGeng, Angela S. Lin, Tianhe Yu, and Alexei A. Efros. Real-time user-guided image colorization with learned deep pri-ors.ACM TOG (Proc. SIGGRAPH), 36(4):119:1–119:11,2017.

[20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2015.

[21] S. Guadarrama, R. Dahl, D. Bieber, M. Norouzi, J. Shlens, and K. Murphy, "Pix Color: Pixel recursive colorization," 28th British Machine Vision Conference (BMVC), 2017

[22] Aaron van den Oord, NalKalchbrenner, OriolVinyals, LasseEspeholt, Alex Graves, ¨ and KorayKavukcuoglu. Conditional Image Generation with PixelCNN Decoders. arXiv:1606.05328, 2016.

[23] Ryan Dahl, Mohammad Norouzi, and Jonathon Shlens. Pixel recursive super resolution. arXiv:1702.00783, 2017.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[25] R J Williams and D Zipser. A learning algorithm for continually running fully recurrent neural networks. Neural Computation, 1(2), 1989.
[26] K. Nazeri, E. Ng, and M. Ebrahimi, "Image colorization using generative adversarial networks," in International Conference on Articulated Motion and Deformable Objects. Springer, 2018, pp. 85–94.
[27] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, SherjilOzair, Aaron Courville, and YoshuaBengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.
[28] Bolei Zhou, AdityaKhosla, AgataLapedriza, Antonio Torralba, and AudeOliva. Places: An image database for deep scene understanding. 2016.
[29] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. 2014.
[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2015.
[31] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, 2015.
[32] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In Proc. ICML, volume 30, 2013.
[33] Kaiming He, Xiangyu Zhang, ShaoqingRen, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, 2015.
[34] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. 2016.
[35] Alec Radford, Luke Metz, and SoumithChintala. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015.
[36] Tim Salimans, Ian Goodfellow, WojciechZaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In Advances in Neural Information Processing Systems, pages 2234–2242, 2016.
[37] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, BiswaSengupta, and Anil A Bharath. Generative adversarial networks: An overview. 2017.
[38] P. Vitoria, L. Raad, C. Ballester, "ChromaGAN: Adversarial Picture Colorization with Semantic Class Distribution", Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2445-2454, 2020.
[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," International Conference on Learning Representations, 2015.
[40] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1125–1134, 2017.
[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.
[42] P. Hensman, K. Aizawa, "cGAN-based Manga Colorization Using a Single Training Image", IEEE Kyoto, Japan, 2018.
[43] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and KiyoharuAizawa, Sketch-based manga retrieval using manga109 dataset, Multimedia Tools and Applications, Springer, pp.1-28, 2016.
[44] Manga 109,2015 http://www.manga109.org
[45] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and KiyoharuAizawa, Sketch-based manga retrieval using manga109 dataset, Multimedia Tools and Applications, Springer, pp.1-28, 2016.
[46] Morevna Project, 2016, https://morevnaproject.org/
[47] Kota Ito, Yusuke Matsui, Toshihiko Yamasaki, and KiyoharuAizawa, Separation of Manga Line Drawings and Screentones, Eurographics, pp.78-76, 2015.
[48] Song-Hai Zhang, Tao Chen, Yi-Fei Zhang, Shi-Min Hu, and Ralph R Martin, Vectorizing cartoon animations, IEEE Transactions on Visualization and Computer Graphics, vol.15, no.4, pp.618-629, 2009.
[49] T. Fontanini, E. Iotti, A. Prati, " MetalGAN: a Cluster-based Adaptive Training for Few-Shot Adversarial Colorization" , International Conference on Image Analysis and Processing, ICIAP 2019: Image Analysis and Processing – ICIAP 2019 pp 280-291.
[50] 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. pp. 1125–1134 (2017).