Black and White Photos Colorization

Md Akibul Islam Alvee

Department of Computer Science & Engineering
Shahjalal University of Science & Technology, Sylhet

March 10, 2022

Contents

Abstract

Colorization of a black and white image is a challenging topic of ongoing research in Computer Science.Image colorization is the process of taking an input black and white image and then producing an output colorized image that represents the semantic colors and tones of the input.Previously, a lot of research work was conducted on this topic.There were different methods, but unfortunately all of them needed human annotation. Deep neural network based colorization is the first computer automated colorization which doesn't need human interaction while colorization. We reviewed existing research work on this topic,then we implemented one of the most recent works which is proposed in a paper entitled as 'ChromaGAN: Adversarial Picture Colorization with Semantic Class Distribution' . This is an adversarial learning colorization approach coupled with semantic information. With the semantic clues of the image, the generative network predicts the chromaticity of that image. This model is trained via a fully supervised strategy. Later, qualitative and quantitative results show the capacity of the proposed method to colorize images in a realistic way achieving state-of-the-art results.

1 Introduction

The aim of image colorization is to add colors to a gray image such that the colorized image is perceptually meaningful and visually appealing. The problem is ill-conditioned and inherently ambiguous since there are potentially many colors that can be assigned to the gray pixels of an input image. Hence, there is no unique correct solution and human intervention often plays an important role in the colorization process.Before the emergence of deep learning techniques, the most effective methods relied on human intervention, usually through either user-provided color scribbles or a color reference image.Currently, digital colorization of black and white visual data is a crucial task in areas so diverse as advertising and film industries, photography technologies or artist assistance. Although important progress has been achieved in this field, automatic image colorization still remains a challenge. In recent years, convolutional neural network (CNN) have emerged as the de facto standard for solving image classification problems, achieving error rates lower than 4% in the ImageNet challenge. CNNs owe much of their success to their ability to learn and discern colors, patterns, and shapes within images and associate them with object classes.

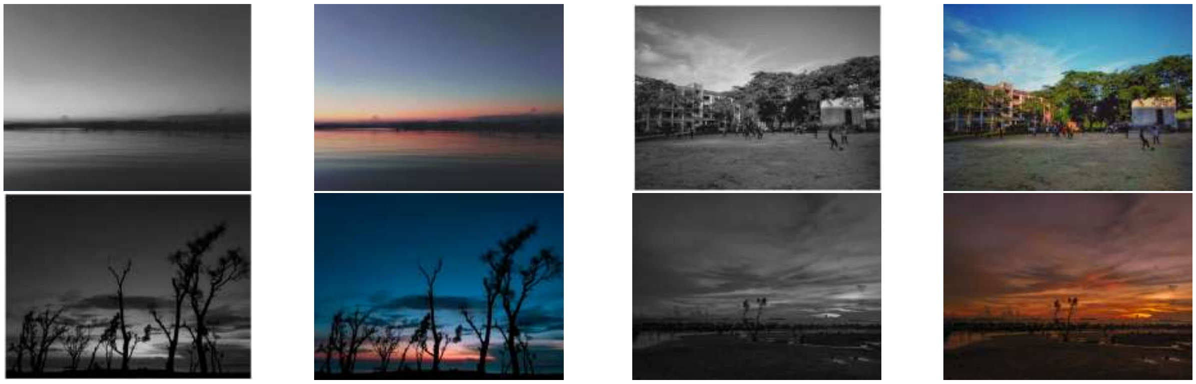

Figure 1: Comparison between colored photo and BnW photo

The paper we've selected, they proposed a fully automatic end-to-end adversarial approach called ChromaGAN. It combines the strength of generative adversarial networks (GANs) with semantic class distribution learning.ChromaGAN shows variability by colorizing differently some objects belonging to the same category that may have several real colors.
The contributions of this work include:

- An adversarial learning approach coupled with semantic information leading to a three term loss combining color and perceptual information with semantic class distribution.
- An unsupervised semantic class distribution learning.
- A perceptual study showing that semantic clues coupled to an adversarial approach yields high quality results.

2 Previously Used Methods

So far,the colorization techniques that have been proposed can be classified in three classes:Scribble-based, Exemplar-based and Deep learning-based methods. The first two classes depend on human intervention. The third one is based on learning leveraging the possibility of easily creating training data from any color image.

2.1 Scribble Based Colorization

The user provides local hints, as for instance color scribbles, which are then propagated to the whole image.There are two parts,segmentation and filling.The error ratio in segmentation remains to be high, which means that a lot of user-interventions are needed to fix the errors, making colorization a tedious, time-consuming and expensive task.Also,these methods suffer from requiring large amounts of user inputs in particular when dealing with complex textures. Moreover,choosing the appropriate color is not an easy task.Though ,a lot of improvement has been done in this sector.But the processing time is still too long, and users still need to be very careful when choosing the colors and strokes positions.

2.2 Exemplar Based Colorization

This method transfers the color information of a reference image to a grayscale image.Here the user provides a suitable image,using the pixel intensity and neighborhood statistics to find a similar pixel in the reference image and then transfer the color of the matched pixel to the target pixel. But finding a suitable reference image is tough,to sort out this extra burden,reference image will be provided by the internet.There is lack of spatial coherency which yields unsatisfactory results. Although this type of methods reduce significantly the user inputs, they are still highly dependent on the reference image which must be similar to the grayscale image.

2.3 Deep Learning Based Colorization

A fully-automatic colorization method formulated as a least square minimization problem solved with deep neural networks.This is basically fully automatic methods where many recent works have been done.These approaches are improved by the use of CNNs and large-scale datasets.However, these approaches aim to produce a single plausible result, even though colorization is intrinsically an ill posed problem with multi-model uncertainty.Colorization is done by globally biasing the hue, or by matching global statistics to a target histogram. Within the training phase, the semantic class label is used to assist in the learning of the global image feature.

3 Proposed Approach

Here,the input image is a grayscale image,the model needs to learn a mapping $\mathscr{G} : L \rightarrow (a, b)$ such that I=(L,a,b) is a plausible(having geometric, perceptual and semantic photo-realism) image and a,b are the chrominance channel images in CIE Lab color space.CIE Lab color space is a 3D color space,which has two parts. One is the grayscale L axis,which means the Luminosity and other (a,b) refers to the chrominance channel.In this approach,mapping $\mathscr{G}$ is learnt by the adversarial leaning strategy.Here,generator predicts the chrominance channel (a,b) and in parallel discriminator evaluates how realistic the colorization is I=(L,a,b) of L.The adversarial energy learns the parameters θ and ω of the generator $\mathscr{G}_\theta$ and the discriminator $\mathscr{D}_\omega$ respectively.Here the generator $\mathscr{G}_\theta$ will not only learn to generate color but also a class distribution vector, denoted by, $y \in \mathbb{R}^m$, where m is the fixed number of classes. This provides information about the probability distribution of the semantic content and objects present in the image.The generator model combines two different modules. Denoted by $\mathscr{G}_\theta = (\mathscr{G}^1_{\theta_1}, \mathscr{G}^2_{\theta_2})$, where $\theta = (\theta_1; \theta_2)$ stand for all the generator parameters, $\mathscr{G}^1_{\theta_1} : L \rightarrow (a, b)$ and $\mathscr{G}^2_{\theta_2} : L \rightarrow y$.

3.1 The Objective Function

The objective loss is defined by

$$\mathscr{L}(\mathscr{G}_\theta, \mathscr{D}_\omega) = \mathscr{L}_e(\mathscr{G}^1_{\theta_1}) + \lambda_g \mathscr{L}_g(\mathscr{G}^1_{\theta_1}, \mathscr{D}_\omega) + \lambda_s \mathscr{L}_s(\mathscr{G}^2_{\theta_2}) \tag{1}$$

The first term denotes the color error loss

$$\mathscr{L}_e(\mathscr{G}^1_{\theta_1}) = \mathbb{E}_{(L,a_r,b_r)\sim\mathbb{P}_r}[||\mathscr{G}^1_{\theta_1}(L) - (a_r, b_r)||^2_2] \tag{2}$$

where $\mathbb{P}_r$ stands for the distribution of real color images and $||.||_2$ for the Euclidean norm. Notice that Euclidean distance in the Lab color space is more adapted to perceptual color differences.Then

$$\mathscr{L}_s(\mathscr{G}^2_{\theta_2}) = \mathbb{E}_{L\sim\mathbb{P}_{rg}}[KL(y_v||\mathscr{G}^2_{\theta_2}(L)||] \tag{3}$$

denotes the class distribution loss, where $\mathbb{P}_{rg}$ denotes the distribution of grayscale input images, and $y_v \in \mathbb{R}^m$ the output distribution vector of a pre-trained VGG-16 model applied to the grayscale image. KL stands for the Kullback-Leibler divergence.

Finally, $\mathscr{L}_g$ denotes the WGAN loss which consists of an adversarial Wasserstein GAN loss.Instead of other GAN losses,WGAN favours nice properties such as avoiding vanishing gradients and mode collapse, and achieves more stable training.To compute it,they use the Kantorovich-Rubinstein duality.They also include a gradient penalty term constraining the L_2 norm of the gradient of the discriminator with respect to its input and, thus, imposing that $\mathscr{D}_\omega \in \mathscr{D}$,where $\mathscr{D}$ denotes the set of 1-Lipschitz functions.To sum up, the WGAN loss is defined by

$$\mathscr{L}_g(\mathscr{G}^1_{\theta_1}, \mathscr{D}_\omega) = \mathbb{E}_{\widetilde{I}\sim\mathbb{P}_r}[\mathscr{D}_\omega(\widetilde{I})] - \mathbb{E}_{(a,b)\sim\mathbb{P}_{\mathscr{G}^1_{\theta_1}}}[\mathscr{D}_\omega(L,a,b)] - \mathbb{E}_{\widehat{I}\sim\mathbb{P}_{\widehat{I}}}[(||\nabla_{\widehat{I}}\mathscr{D}_\omega(\widehat{I})||_2-1)^2] \tag{4}$$

where $\mathbb{P}_{\mathscr{G}^1_{\theta_1}}$ is the model distribution of $\mathscr{G}^1_{\theta_1}(L)$, with $L \sim \mathbb{P}_{rg}.\mathbb{P}_{\widehat{I}}$ is implicitly defined sampling uniformly along straight lines between pairs of point sampled from the data distribution $\mathbb{P}_r$ and the generator distribution $\mathbb{P}_{\mathscr{G}^1_{\theta_1}}$. The minus before the gradient penalty term in (4) corresponds to the fact that, in practice, when optimizing with respect to the discriminator parameters, this algorithm minimizes the negative of the loss instead of maximizing it. From the previous loss (1),they compute the weights of $\mathscr{G}_\theta$;$\mathscr{D}_\omega$ by solving the following min-max problem.

$$min(\mathscr{G}_\theta)max(\mathscr{D} \in \mathscr{D}_\omega) \qquad \mathscr{L}(\mathscr{G}_\theta, \mathscr{D}_\omega) \tag{5}$$

The hyperparameters λ_g and λ_s are fixed and set to 0.1 and 0.003, respectively.

The Adversarial Strategy and The GAN Loss

The min-max problem (5) follows the usual generative adversarial game between the generator and the discriminator networks. The goal is to learn the parameters of the generator so that the probability distribution of the generated data approaches the one of the real data, while the discriminator aims to distinguish between them.The proposed Lg loss favours perceptually real results.The adversarial GAN model produces sharp and colorful images favouring the emergence of a perceptually real palette of colors instead of ochreish outputs produced by colorization using only terms such as the *L2* or *L1* color error loss.

Color Error Loss and Class Distribution Loss

They chose to learn two chrominance values (a; b) per-pixel using the *L2* norm.Only using this type of loss yields ochreish outputs. However, the perceptual GAN-based loss relaxes this effect making it sufficient to obtain notable results.Again,the VGG-16 model was trained on color images; in order to use it without any further training,they re-shape the grayscale image as *(L; L;L)*. The class distribution loss adds semantic interpretation of the scene.

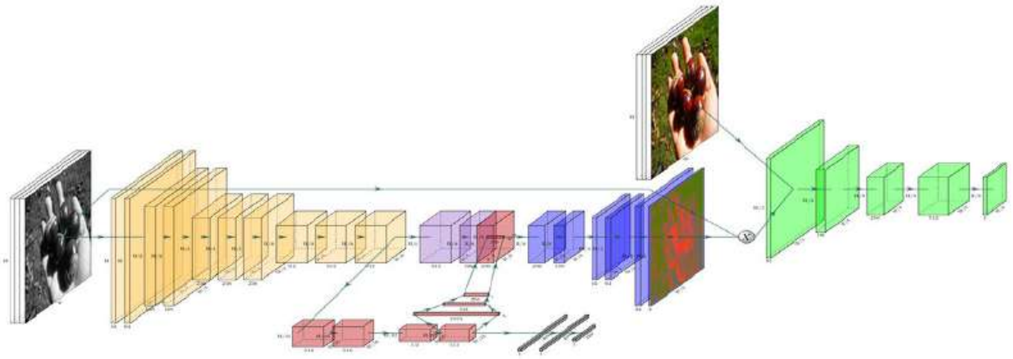

Figure 2: Overview of ChromaGAN, able to automatically colorize grayscale images. It combines a Discriminator network, $\mathscr{D}_\omega$ (in green), and a Generator network, $\mathscr{G}_\theta$

4 Detailed Model Architecture

The proposed GAN architecture contains three distinct parts. The first two, belonging to the generator,focus on geometrically and semantically generating a color image information (a; b) and classifying its semantic content. The third one belongs to the discriminator network learning to distinguish between real and fake data.

Generator Architecture

The generator $\mathscr{G}_\theta$ is divided in three stages.In first two stage, the generator network($\mathscr{G}_\theta$) is divided into two sub-network denoted by ($\mathscr{G}^1_{\theta_1}$,$\mathscr{G}^2_{\theta_2}$),both of them will take as input a grayscale image of size $H\times W$. The subnetwork $\mathscr{G}^1_{\theta_1}$ outputs the chrominance information, (a; b) = $\mathscr{G}^1_{\theta_1}$*(L)*, and the subnetwork $\mathscr{G}^2_{\theta_2}$ outputs the computed class distribution vector, y = $\mathscr{G}^2_{\theta_2}$*(L)*.The first subnetwork (in purple in Fig. 2)process the data by using two modules of the form Conv- BatchNorm-ReLu. The second subnetwork (in red), first processes the data by using four modules of the form Conv-BatchNorm-ReLu, followed by three fully connected layers(in red). This second path (in gray) outputs $\mathscr{G}^2_{\theta_2}$ providing the class distribution vector. To generate the probability distribution y of the m semantic classes, they apply a softmax function.In the third stage both branches are fused (in red and purple in Fig. 2) by concatenating the output features.

Discriminator Architecture

The discriminator network $\mathscr{D}_\omega$ is based on the Markovian discriminator architecture.In order to model the high frequencies, the PatchGAN discriminator focuses on local patches rather than giving a single output for the full image, it classifies each patch as real or fake.

5 Implementations Details

We found the existing code for this method,but there was version mismatching.So,we rewrite the existing code,which is: Coloring Black and White Photos.This model is trained with 1.3M images taken from Image-Net which contains object from different 1000 categories of different color conditions.Each images of the training images is resized to 224 × 224.This model is trained with 5 epochs,and the set of each batch size is 10.We collected the pretrained weight from the existing GitHub repository of this paper.And it takes almost 4.4 milliseconds for the prediction of colorization.

6 Results

We tested some photos with this model.Firstly,the model converts it into black and white photo,then colorizes it.Now If we see some output generated by this model.

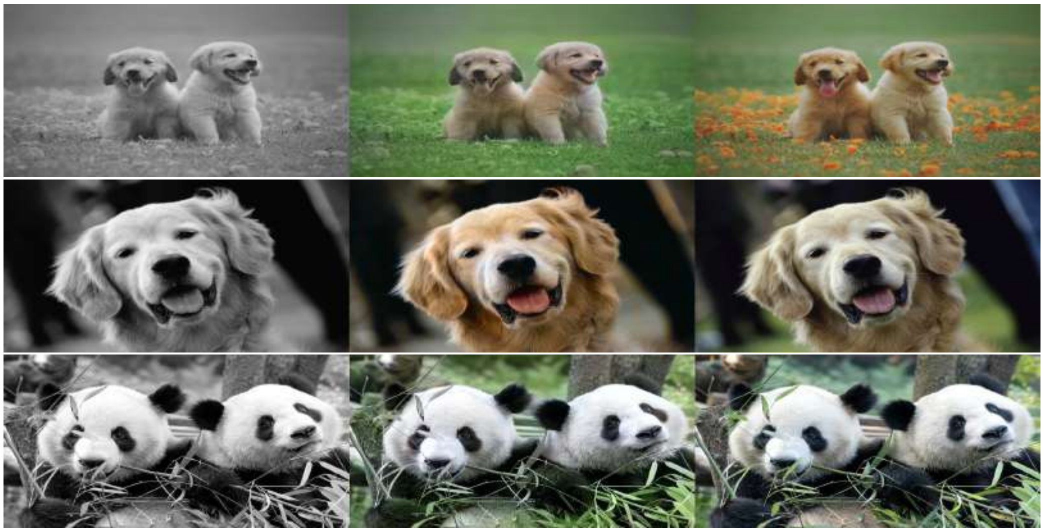

Figure 3: Result using ChromaGAN,from left to right:Gray scale,ChromaGAN,Real image

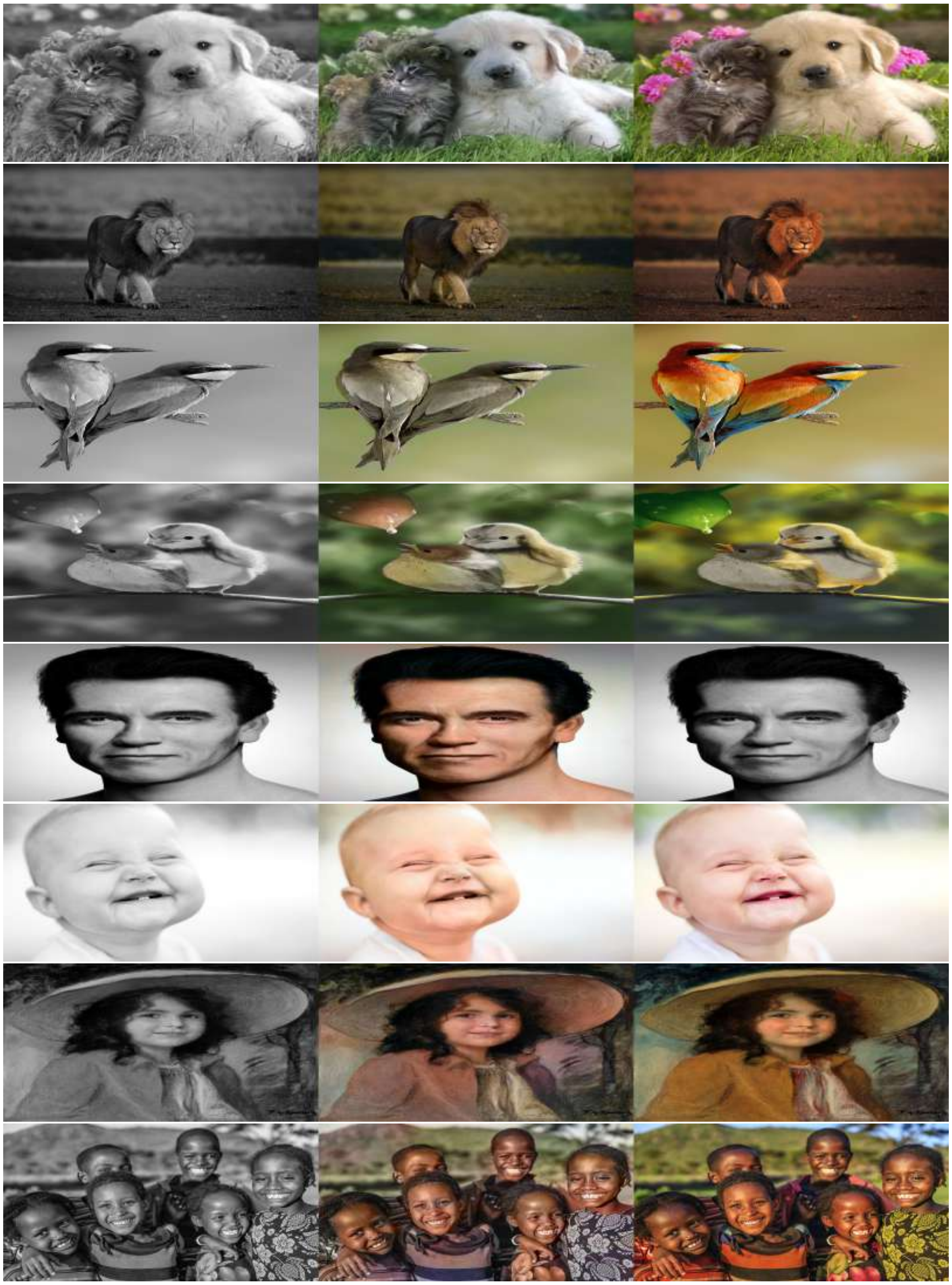

Figure 4: Result using ChromaGAN,from left to right:Gray scale,ChromaGAN,Real image

7 Comparison

If we compare this model with other Neural Network based models,it is seen that it works comparatively better than those in accuracy and PSNR value .The PSNR block computes the peak signal-to-noise ratio, in decibels, between two images. This ratio is used as a quality measurement between the original and a compressed image. The higher the PSNR, the better the quality of the compressed, or reconstructed image.

Method	Naturalness
Real images	87.1%
ChromaGAN	76.9%
ChromaGAN w/o class	70.9%
ChromaNet	61.4%
Iizuka et al. [2016]	53.9%
Isola et al. [2017]	27.6%
Larsson et al. [2016]	53.6%
Zhang et al. [2016]	52.2%

Table 1: The values shows the mean naturalness over all the experiments of each method.

Method	PSNR(dB)
ChromaGAN	24.98
ChromaGAN w/o class	25.04
ChromaNet	25.57
Iizuka et al. [2016]	23.69
Isola et al. [2017]	21.57
Larsson et al. [2016]	24.93
Zhang et al. [2016]	22.04

Table 2: Comparison of the average PSNR values for automatic methods.

8 Limitations

There are some limitations in this model.Outputs do not satisfy properly if we use unconventional photos.We try to collect some photos,which are not familiar to this model,then we check this model with these photos. As this model is semantic based or content based,the more it is familiar to the content,the better it performs.

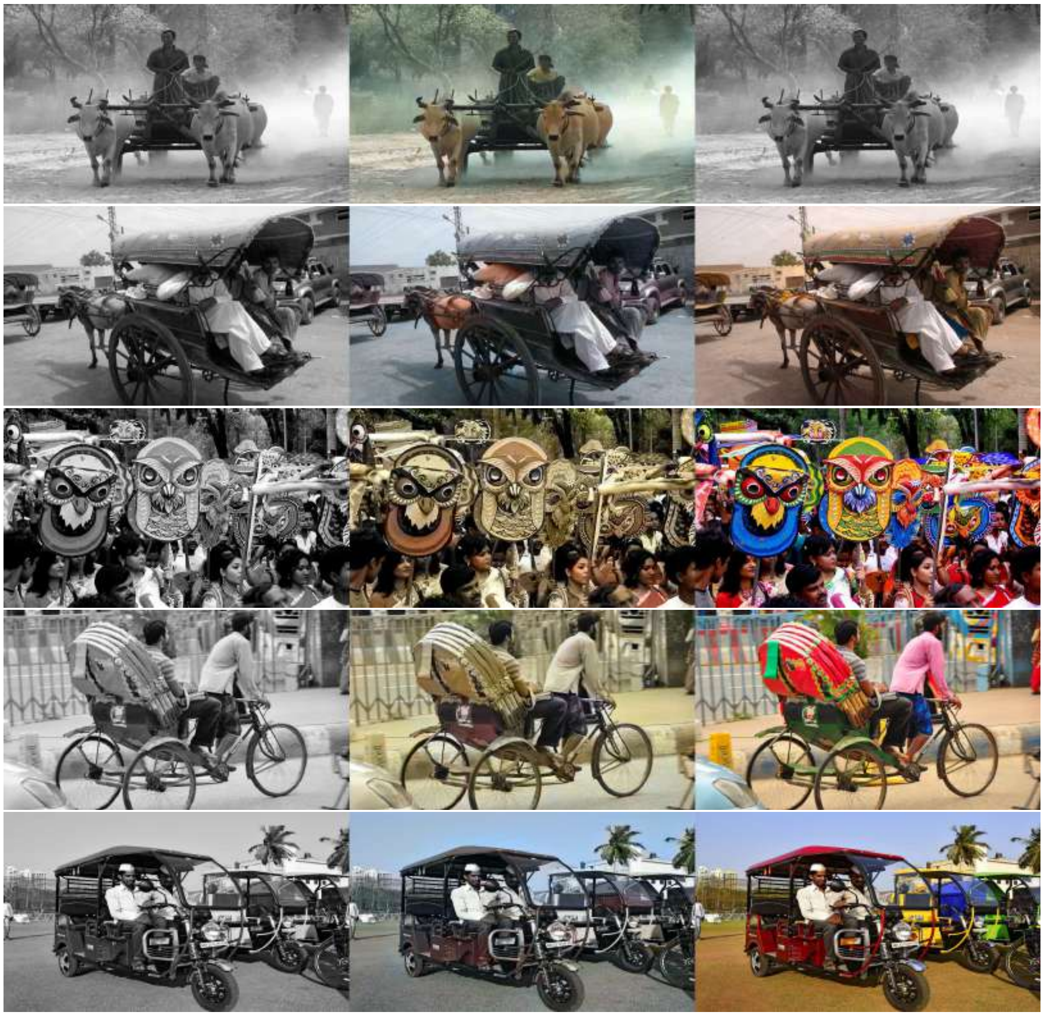

Figure 5: Result using ChromaGAN,from left to right:Gray scale,ChromaGAN,Real image

9 Conclusion

Among a lot of colorization we choose one of the recent work that has been conducted on this topic,though we reviewed all related articles and paper on this topic such as Arjovsky et al. [2017] He et al. [2018] Iizuka et al. [2016] Isola et al. [2017] Karras et al. [2019] Koo [2016] Larsson et al. [2016] Simonyan and Zisserman [2014] Zhang et al. [2016] Bao and Fu [2019] Varga and Szirányi [2016].Then we selected this paper to implement.This model is based on adversarial strategy that captures geometric, perceptual and semantic information.It is seen that this model outperforms all other existing neural network based models in terms of accuracy and in PSNR values. But the door is always open for doing better.Further,specifically focused on one topic,this model can yield better and pleasing output.

References

M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

B. Bao and H. Fu. Scribble-based colorization for creating smooth-shaded vector graphics. *Computers & Graphics*, 81:73–81, 2019.

M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan. Deep exemplar-based colorization. *ACM Transactions on Graphics (TOG)*, 37(4):1–16, 2018.

S. Iizuka, E. Simo-Serra, and 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 (ToG)*, 35(4):1–11, 2016.

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.

T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

S. Koo. Automatic colorization with deep convolutional generative adversarial networks. *CS231n*, 2016.

G. Larsson, M. Maire, and G. Shakhnarovich. Learning representations for automatic colorization. In *European conference on computer vision*, pages 577–593. Springer, 2016.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

D. Varga and T. Szirányi. Fully automatic image colorization based on convolutional neural network. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 3691–3696. IEEE, 2016.

R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.