International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

Image Colorization using Self attention GAN

Amey Narkhede[1], Rishikesh Mahajan[2], Priya Pacharane[3]

[1,2,3]*Student, Dept. of Computer Engineering, Pillai HOC College of Engineering and Technology, Maharashtra, India*

Abstract - *The problem of coloring of grayscale images is highly ill posed due to large degrees of freedom available while assigning colors. In this paper, we use Deep Learning based Generative Adversarial Networks to produce one of the plausible colorization of black and white images. Many of the recent developments are based on Convolutional Neural Networks(CNNs) which required large training dataset and more resources. In this project, we attempt to extend GAN using self-attention capability to colorize black and white images that combines both global priors and local image features. We also suggest efficient training strategies for better performance. This model works on different types of images including grayscale images and black and white images which are hundred years old.*

Key Words: Image Colorization, Generative Adversarial Networks, Self attention

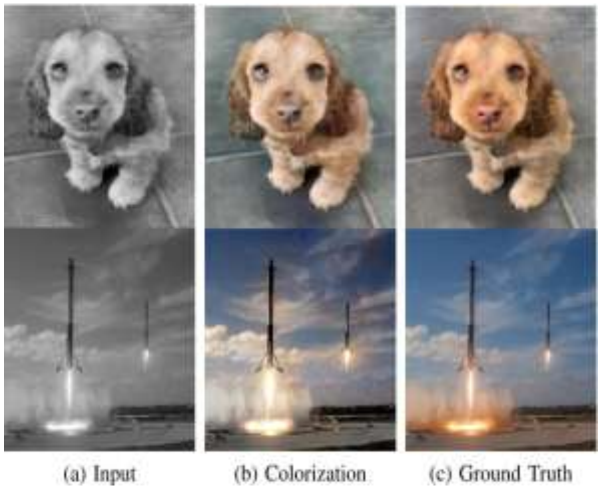

Fig -1: Colorization example. (a) Input grayscale image obtained from (c). (b) Output colorized image. (c) Original color image

1. INTRODUCTION

Coloring grayscale images automatically has been an active area of research in machine learning for a long period of time. This is due to the wide range of applications such color restoration and image colorization for animations. The aim of automatic image colorization technology is to automatically color grayscale images without manual intervention. In this paper, our aim is not necessarily to produce the true ground truth color, but rather to create a plausible colorization of grayscale image that could potentially trick a human observer. As a result, out task becomes more feasible:to model sufficient statistical dependencies between texture of grayscale images and their color versions to produce visually compelling results.

Welsh et al[1] proposed an algorithm that colorized images through texture synthesis. Colorization of grayscale image was carried out by matching the luminance and texture information between the existing color image and the grayscale image to be colorized. However, this proposed algorithm was described as a forward problem, hence all solutions were deterministic. Traditional methods[1],[2],[3] use reference image to transfer color information to target grayscale image. Reinhard et al[4] Tai et al[5] proposed the colorization models based on color transfer by calculating statics in both input and reference image and constricting mapping that map color distribution of reference image to input grayscale image. However, getting reliable reference image is not possible for all cases. Some previous works have trained convolutional neural networks (CNNs) to predict color on large datasets [6],[7]. Cheng et al[6] proposed first fully automatic colorization method conceived as a least square minimization problem solved by deep neural networks. A semantic feature descriptor is presented and supplied as an input to the network. The model extracts various features and colors the input images in small patches. However, the results from these previous attempts tend to look desaturated. One explanation is that [6],[7] use loss functions that encourage conservative predictions. These loss functions are derived from standard regression problems, where the aim is to minimize Euclidean distance between an estimate and the ground truth. Iizuka el al [8] use two stream CNN architecture in which global and local features are fused. Zhang el al [9] use VGG-style single stream network with additional depth and dilated convolutions to map grayscale input to quantaized color value distribution of output.

In this paper, we will explore the method of colorization using Self-Attention Generative Adversarial Networks (GANs) proposed by Goodfellow et al [11] which tries to model the novel loss function. The model is trained on the ImageNet [17] dataset. The goal of image colorization is to append colors to a grayscale image such that the colorized image is perceptually meaningful and visually appealing. This problem is immanently uncertain since there are potentially many colors that can be given to the gray pixels of an grayscale input image (e.g., hair may be colored in black, brown, or blonde). Therefore, there is no unique perfect solution and human involvement often plays an important role in the colorization process. We have proposed ColorGAN based on Self Attention GAN [11] for image colorization task where lightweight U-Net is used as generator. Our proposed model is trained in unsupervised fashion without strictly paired data, which minimizes effort required for data collection.

2. PROPOSED METHOD

2.1 Network Architecture

Self-Attention Generative Adversarial Networks(SAGAN)[11] introduced self attention mechanism into conventional GANs[10]. Self attention mechanism helps create a balance between efficiency and long-range dependencies. We proposed SAGAN[11] based approach and modified it to suit our goal.

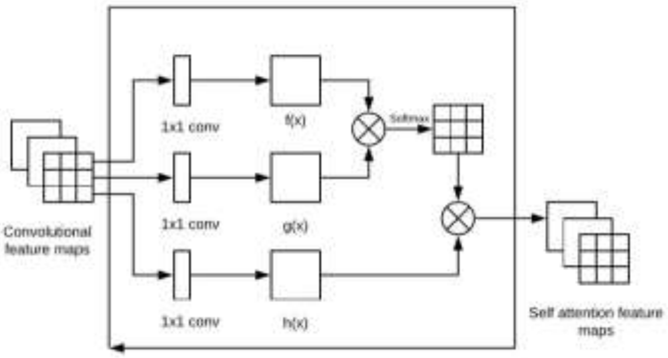

Fig -2: Self attention architecture

The generator network *G* is used to map input grayscale images to the color images. Specifically Multi-layers convolutional network U-net[12] is used as generator. The U-net[12] is modified to incorporate spectral normalization [13] and self attention mechanism described by SAGAN [11]

The discriminator network *D* is just a binary classifier as the task is relatively simple. The generator images from generator is given as input to discriminator. We can also simply use a pre-trained discriminator used for another image-to-image task(like CycleGAN [15]) and fine-tune it. We pre-train both generator and discriminator separately before training both networks in GAN setting using Two Time-Scale Update Rule proposed by Heusel et al [16].

2.2 Methodology

We train generator and discriminator separately before moving towards GAN training. Specifically the generator is represented by $G(z;\Theta_G)$ where z is uniformly distributed noise that is passed as input to generator. The discriminator is represented by $D(x;\theta_D)$ where x is a color image. It gives the output between 0 and 1 which denotes the probability of image belonging to training data.

We use U-net [12] which is modified to include self attention mechanism described by SAGAN [11] and spectral normalization [13]. Initially we train the generator using perceptual/feature loss proposed by Johnson et al [14] and the discriminator as basic binary classifier separately. The perceptual loss [14] is used to just bias the generator to reproduce input image.

Then we train generator and discriminator in GAN setting using Two Time-Scale Update Rule [16]. After small window of time it is observed that discriminator no longer provides constructive feedback to generator. Past this point the quality of generated image oscillates between best you can get at previously described point and bad(distorted colors). It appears to be that there is no productive training after this point.

We can repeat pre-training the discriminator after the initial GAN training and then repeat the GAN training as previously described in the same manner. This results in images with brighter colors but we need to adjust render resolution of generator as output of generator starts to become inconsistent.

3. RESULT

We trained generator and discriminator on ImageNet [17] dataset. We used Adam as optimizer. The initial learning rate for both generator *G* and discriminator *D* are set to 0.0002. The model was implemented in PyTorch framework and trained on Nvidia GPU with 11GB of memory.We observed that in Fig 2. after the inflection point the image quality oscillates between best and bad.

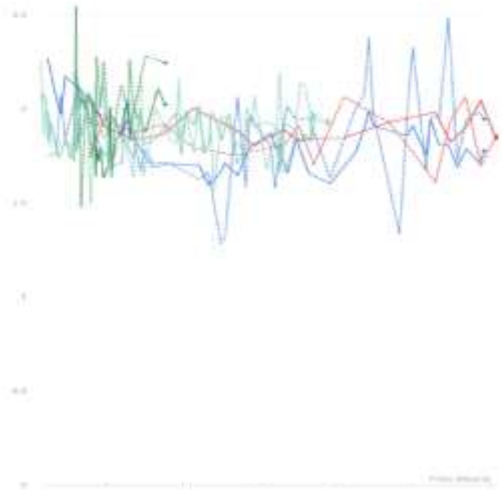

Fig -3: The loss oscillates between high and low.

(a) Input (b) CIC [9] (c) LC [8] (d) Proposed (e) Origin

Fig -4: Our proposed model performs better than previous approaches.

3.1 Comparison against several existing methods

We compared our results with several existing CNN based models [8],[9].

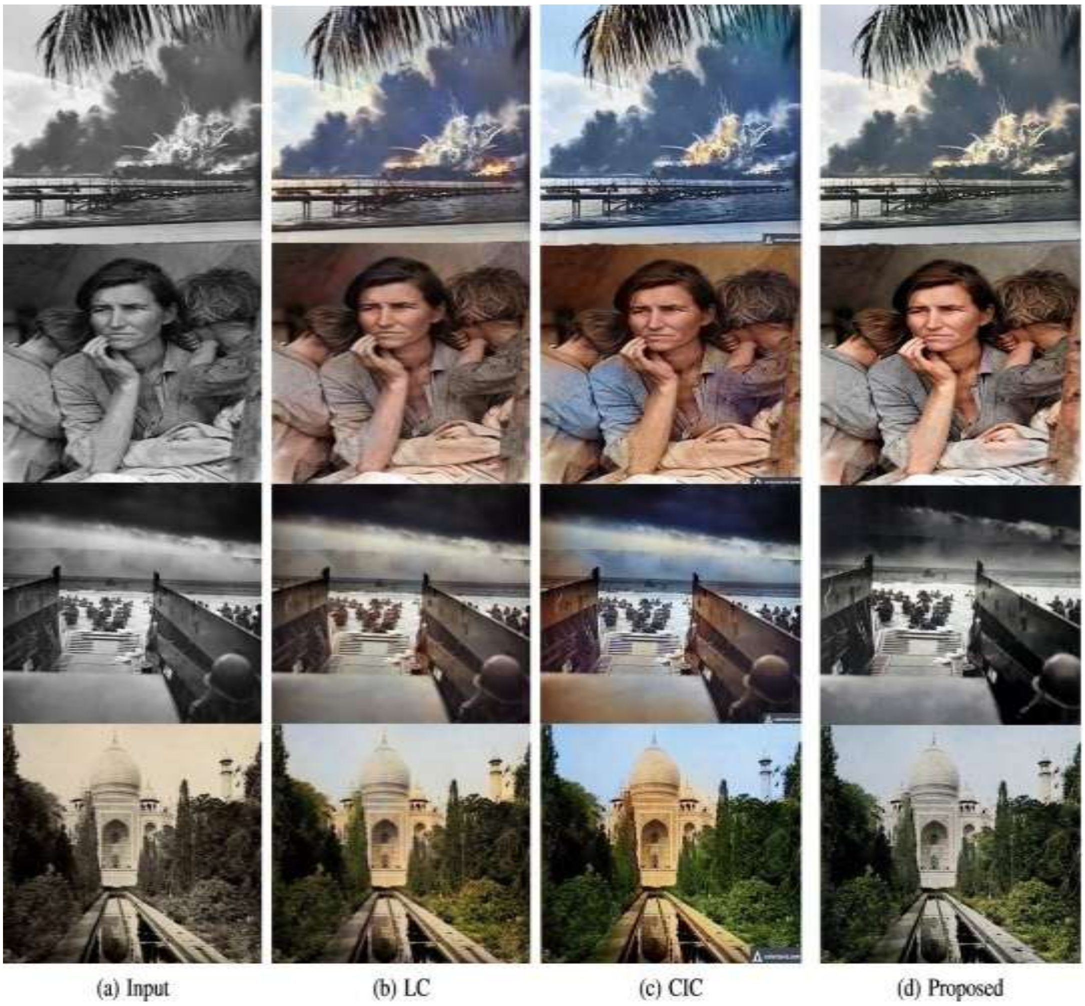

Fig -5: (a) Original grayscale images. (b)[8]has achieved good results using custom CNN and paired data but in some cases in produces distorted colors. (c) CIC [9] produces some unreasonable color artifacts in many cases. (d) Our proposed model circumvents color artifacts and produces reasonable color images.

3.2 Human subjective evaluation

We performed human subjective comparison between different models. The test subject was asked which image look more reasonable and realistic. The final result as shown in fig(comparison results) show that our model can produce more realistic and reasonable coloring. It should be noted that we are aiming to produce one of the many plausible coloring instead of restoring exact original color.

4. CONCLUSION

We have shown that image colorization with self attention GAN come closer to produce more plausible and realistic color photos. We also proposed a new GAN training method to shorten the effective training time. Our model learns representations that is also useful for tasks other than image colorization such that classification and segmentation. This work can be extended to video colorization.

REFERENCES

[1] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller, “Transferring color to greyscale images”, SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniquesJuly 2002 Pages 277–280.

[2] Raj Kumar Gupta, Alex Yong-Sang Chia, Deepu Rajan, Ee Sin Ng, and Huang Zhiyong, “Image colorization using similar images”, Proceedings of the 20th ACM International Conference on Multimedia, 2012.

[3] Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, and Lu Yuan, “Deep exemplar-based colorization”, ACM Transactions on Graphics, vol. 37, pp. 1–16, 07 2018.

[4] Reinhard E., Ashikhmin M., Gooch B., and Shirley, P., “Color transfer between images”, IEEE Computer Graphics and Applications 21, 5 (sep) 2001, 34–41.

[5] Tai Y., Jia J., and Tang C

[6] Cheng, Z., Yang, Q., Sheng, B, "Deep colorization", In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 415–423

[7] Dahl, R, "Automatic colorization.", http://tinyclouds.org/colorize/. (2016)

[8] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi 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, vol. 35, no. 4, pp.1–11, 2016.

[9] Zhang, Richard and Isola, Phillip and Efros, Alexei A, "Colorful Image Colorization", European Conference on Computer Vision, 2016.

[10] Goodfellow, Ian and Pouget-Abadie, Jean and Mirza, Mehdi and Xu, Bing and Warde-Farley, David and Ozair, Sherjil and Courville, Aaron and Bengio, Yoshua, "Generative Adversarial Nets", Advances in Neural Information Processing Systems 27, NIPS, Pages 2672-2680, 2014.

[11] 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.

[12] Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas, "U-net: Convolutional networks for biomedical image segmentation", International Conference on Medical image computing and computer-assisted intervention,234--241, 2015, Springer.

[13] Takeru Miyato and Toshiki Kataoka and Masanori Koyama and Yuichi Yoshida, "Spectral Normalization for Generative Adversarial Networks", International Conference on Learning Representations, 2018.

[14] Johnson, Justin and Alahi, Alexandre and Fei-Fei, Li, "Perceptual losses for real-time style transfer and super-resolution", European Conference on Computer Vision, 2016.

[15] Zhu, Jun-Yan and Park, Taesung and Isola, Phillip and Efros, Alexei A, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", Computer Vision (ICCV), 2017 IEEE International Conference on Computer Vision, 2017.

[16] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", Advances in Neural Information Processing Systems 30 (NIPS 2017).

[17] Deng, J. and Dong, W. and Socher, R. and Li, L.-J. and Li, K. and Fei-Fei, L., "ImageNet: A Large-Scale Hierarchical Image Database", Conference on Computer Vision and Pattern Recognition, 2009.