International Journal for Modern Trends in Science and Technology, 8(S03):225-230, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST08S0346
Available online at: https://www.ijmtst.com/vol8si03.html

Performance Analysis of Generating Cartoon like Images from Realistic Images Using Various Cycle GAN Techniques

Kanchi Ramya, Dr.N.Leelavathy

Department of Computer Science and Engineering, Godavari Institute of Engineering and Technology (A), JNTUK, Kakinada.

To Cite this Article
Kanchi Ramya and Dr.N.Leelavathy. Performance Analysis of Generating Cartoon like Images from Realistic Images Using Various Cycle GAN Techniques. International Journal for Modern Trends in Science and Technology 2022, 8(S03), pp. 225-230. https://doi.org/10.46501/IJMTST08S0346

Article Info
Received: 26 April 2022; Accepted: 24 May 2022; Published: 30 May 2022.

ABSTRACT

In this paper, we present performance study of cartoon-like picture generation from actual photos utilizing several GAN methods. The idea of cartooning is to create cartoonic representations of actual input pictures. Three picture white-box displays are produced:1. Surface Depiction-It contains a smooth surface of cartoon images.2. Structure Depiction- It refers to the small color blocks and flat global information in the workflow style of celluloids.3. Texture Depiction- In cartoon pictures it displays high-frequency texture, curves and features. The cartoonic image outcomes on supplied pictures are suggested by utilizing Generative Adversarial Network (GAN) methods. GAN frame is used to learn the extracted pictures and cartoonize the realistic images provided.These photos may be used for movie animations, fun activities, online avatars, social images, etc. These pictures are recommended ifsomeone wants to publish genuine images to prevent the danger of profound counterfeits.The outcomes of both methods should be calibrated to determine the optimum technique.

Keywords: CartoonGAN, Deep Learning, CNN, Tensor flow, UNet, ResNet.

1.INTRODUCTION

A cartoon is a basic diagram of the things we look at in an amusing manner with a lot of bright hues or a cartoon is an image, cursor, or animatronic movie. For youngsters, cartoons are tremendous fun. They are typically found on the small screen or in newspapers, comic books and magazines. Cartoons have solely been forthe notion of entertainment in recent years. But today a day's cartoons are widely utilized for teaching and improving people's perceptions as well as pleasure and enjoyment. In order to create high-quality cartoons, artists must sketch each color zone of the target parts with every line and shade. The prevalent algorithms for picture modification with conventional characteristics cannot generate acceptable cartooning outcomes. Specially developed methods that can automatically convert real world photographs into first-class cartoon images are thus suitable and may save artists time by concentrating on more creative work. Cyclically articulated generative adversarial networks (GANs) are examined to obtain a better grade of Transfer of style, having a unique feature that the model has been taught using unpaid pictures and styled photos. The aim of the cartoon style algorithm is to map photos on a photo multiplex into the cartoon collection while retaining their information. To accomplish this aim, we suggest employing a specialized GAN architecture that can

successfully map pictures using unpaired image sets from realistic photographs to cartoonic images.

- This paper's major contributions are: Use UNet, an end-to-end segmentation technique in which pictures are the same size, input and output to create high quality styled cartoons.
- Use ResNet (Residual Networks), a deep neural network used to build better cartoons as a backbone for various computervision applications.
- Evaluate the performance of each of the aforementioned techniques in terms of number of parameters, then analyses the loss comparison between two ways and provide the best way to improve the loss.

2. LITERATURE REVIEW

The CNNs is a kind of profound neural networks used to assess visual fantasy; they explain numerous computer vision problemsto replace NPR algorithms which require significant effort to accomplish each style transmission. All methods suggested should use a single content picture image style and their results are largely dependent on thechosen image style as the separation of styles and content in the picture style is an inevitable uncertainty. Genetic Advertising Networks (GANs)[5] is a possible alternative approach to the synthesis of pictures that are generative models of deep learning. They are state-of-the-art in many applications, including image conversion, image technology and high-resolution images. The GAN model is built on a system that trains two networks, a generator and a discriminator periodically. The adversarial loss generated by the discriminator transfers the images to the multiple targets. To overcome this problem, Cycle GAN, a system that can interpret pictures utilizing unparalleled training data, has been recently created. Two set of GAN models are being simultaneously trained by Cycle GAN, CLASS Map A by CLASS B and CLASS B by CLASS A. The loss is reflected in the mapping of images to the same class. Our methods combine two sets of real-life photographs and cartoon images to create a cartoon look.

3. PROPOSED SYSTEM

GAN Framework:

The GAN framework consists of two CNNs.G-generator network that is trained to produce output that substitutes the discriminatory network. The next one is Discriminator D, who classifies whether the image is of the actual multiple or synthetic goal. We build the network of the generator and discriminator network to fit theuniqueness of the picture; see Figure 1.

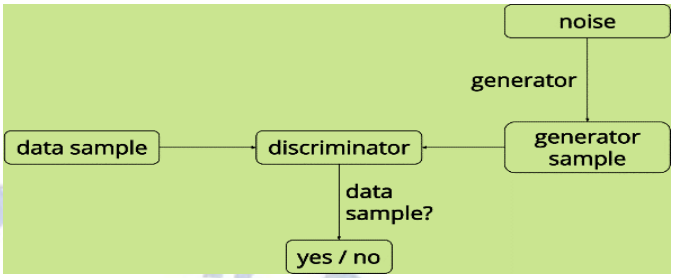

Figure 1**: Overview of Generator anddiscriminator.**

We propose a technique via which real world photos are converted into caricature images as a mapping function, translating multiple P photographs to multiple C caricatures. Data are utilised to learn the mapping function

$S_{data}(p) = \{pi \mid i = 1 \ldots N\} \subset P$ and

$S_{data}(c) = \{ci \mid i = 1 \ldots M\} \subset C$,

where N and M are the numbers of photo and cartoon images in the training set respectively.

The discriminating function D, like earlier GAN frameworks, is intended to compel G to accomplish his purpose by differentiating between incoming pictures in the drawingmanifold from other images and the loss for G. The loss function must be L, the network weight must be D and the network weight[7][8]. The goal here is to solve the problem:

$$(G, D) = \arg\min_{G}\max_{D} L(G, D)$$

The G network generator is used in CartoonGAN to map photos in various diagrams. Once the model is trained, cartoon style is produced. G begins with a flat convolution phase followed by two down blocks to spatially compress the images. In this phase, relevant local signals are retrieved for downstream processing. Eight remaining blocks with a similar arrangement are then used to construct the content and many features. The rest of the block layout is used. Finally, for the resultant images in the cartoon style, two up-coevolutionary blocks containing partly stroke Convolutional layer are constructed.

Figure 2 below shows the generator and discriminator architecture for the proposed GAN cartoon, where k is kernel-size, n is the number of mappings, and s is a step in every revolutionary layer.

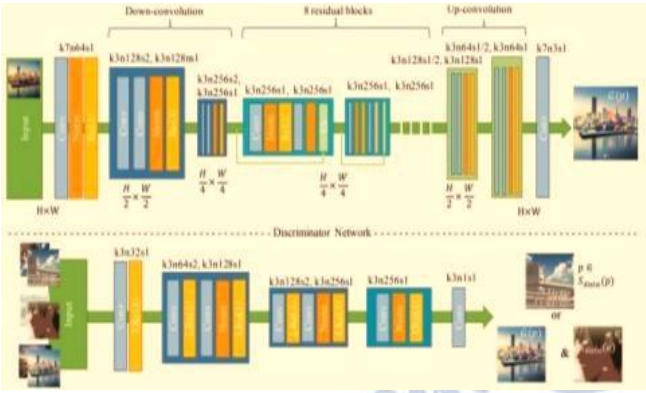

Figure 2: **Overview of Generator and discriminator.**

The Discriminator Network (D) confirms that the image is a true cartoon image, complementing the network generator (G). We use a simple patch level discriminator with fewer parameters in D instead of a typical complete image discriminator to validate if the picture has been a cartoon or not. In contrast to object categorization, discrimination is based on the local features of the image in the cartoon style. The Dnetwork is thus meant to be small. The network provides two phases of Convolutional blocks following the phase with flat layers, reducing the resolution and encoding important local classification features. A component block and a 3 x 3 convolution layer are then used to obtain the classification answer.

UNet Architecture:

The basic concept of the Convolutional Neural Networks (CNN) beach is to learn how to map a picture and alter it so that more detailed drawings are made. This works well with classification issues since the picture becomes a vector that is utilized further for categorization. However, in image segmentation, we need to not onlytransform the feature map into a vector, but also rebuild a n image This vector. From this vector. It's an enormous task since it is harder to turn a vector into an image than vice versa. The entire concept of UNet is around this issue. Use the same characteristic maps used to restrict a vector to a picture that is segmented. This would preserve the structural truthfulness of the image which would greatly reduce alteration.

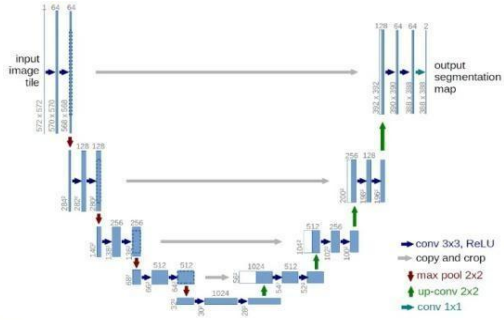

Figure 3: UNET Architecture

Figure 3 shows the construction as a 'U' thatexplains its name. It comprises of three areas: contraction, bottleneck and growth. The initial part is a combination comprising many blocks of contraction. Individual block receives the input of two 3X3 convolution layers and a max 2X2 pooling. The number of maps after each block is doubled so that buildings can efficiently learn about the complicated structures. Between the contraction layer and the expansion layer the lowest layer intermediates. It utilizes two 3X3 CNN layers which are examined by 2 X2 up. In the expansion section lays the heart of UNet design. Like in the contraction layer, it consists of multiple blocks of expansion. Each block is sent to two layers of 3X3 CNN followed by a sample layer of 2X2. And every block of maps utilized by the Convolutional layer is half symmetrical. However, every time the input is additionally added to the corresponding contraction layer via feature maps. This step would ensure that the characteristics learnt when the picture is contracted are utilized to rebuild it. The number of blocks of expansion is the same as the number of blocks. The secondary mapping then goes through another 3X3 CNN layer, with the feature map number equal to the required number of segments.

ResNet Architecture:

Residual neural networks [25] are the kind of neural network that uses identity mapping or is popularly referred as s ResNet. This implies that the input is sent straight to some layer or as a shortcut to another one. Consider Figure 4 below, which displays a fundamental residual block. The most essential idea here is the skip link or shortcut. It can be found. Skip connection is essentially the identity mapping where the previous layer is appended to the other layer's output immediately.

In a ResNet there are two major block kinds, mostly dependent on whether the input/output dimensions are same or different.

Identity Block: The identity block shown in figure 5 below is the typical block in ResNet applications and resembles the one that has the same dimension of input activation as output creation.

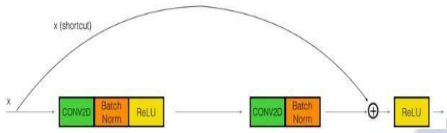

Figure 5: **Identity Block**

Convolutional Block: This kind of block may be used in Figure 5 below if the dimension of the input and output does not match. The difference with the identity block is that the shortcut route contains a CONV2D layer.

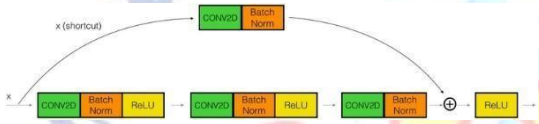

Figure 6**: Convolutional Block**

ResNet utilizes four residual blocks with the same number of output channels in the proposed system. In the meanwhile, a maximum pooling layer with a stride of 2 has already been utilized, the height and breadth should not be reduced. For each of the following modules, the initial residual block doubles the number of channels compared to the previous module and splits height and width into fifty-fifty.

4. EXPERIMENTAL RESULTS

COMPARISON BETWWEN UNET ANDRESNET:

The outcomes of caricatured human faces and cartoon ashes utilizing CNN architecture of UNet and ResNet are discussed below:

The Discriminator model:

Contains Convolutional neural networks (CNNs) and batch layers of standardisation alternating. For the final activation layer, we utilise the Sigmoid activation function and the Leaky ReLU activation function on all levels.

The Generator model:

Contains transposes, alternating with one another of CNNs and batch normalization layers. The Tanh function is used for final activation and the ReLU function for all other layers. Data The training details include pictures and cartoon images from the actual world, while the test data also include photos from the real world and test cartoon faces. All pictures are redimensioned and converted into 128*128 pictures. 1000 photographs from Flickr are downloaded, 500 pictures for training andtesting.

Loss calculation:

For each pixel, UNet and RENET are using a relatively new loss weighting technique that increases our weight on the boundary of segmented objects. First, the pixel-wise activation feature is applied to the resulting picture, followed by the cross-entropy loss feature. Thus, we classify each pixel into one of the classes. The notion is that every pixel has to be in one category even in segmentation, and we simply have to ensure that it does. We have thus simply converted a segmentation issue into a multi grade problem and accomplished this extremely effectively in comparison with conventional loss functions.

Below are the pictures produced for human faces and the graph shows the loss of CycleGAN with the UNet architecture.

Loss of adversary Ladv (G, D) The adverse loss[20] is applied to the G and D networks, which influence the animation process in the G-generator network. The number indicates the degree to which the generator G output picture appears like a cartoon. The antagonistic loss shown in figure 7 below

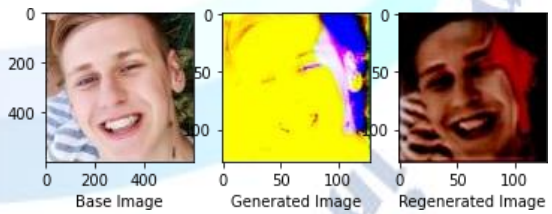

Figure 7: Original photo to generatedimage

Content loss Lcon(G, D)

A further essential objective for the cartoonstylization is to make sure that the resultant cartoon images maintain semantical information from the original pictures. We use the high-level UNET feature maps in CartoonGAN. Below are the pictures for cartoon f aces produced, and the graph shows the CycleGAN loss using UNet architecture. The Loss value generated is 0.2826 Considered for five and ten batch sizes.

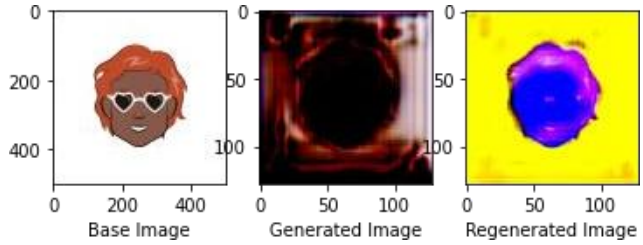

Figure 8: **Original cartoon to redeveloped image after loss**

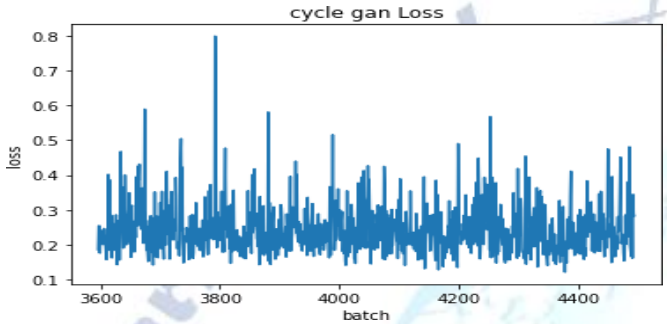

Figure 9: Graph of cycle gan loss

Below are the images for human faces The chart illustrates CycleGAN loss using the **ResNet architecture.** Loss of opponent Ladv Ladv (G, D) The negative losses are applied to both the G and D networks, which affect the G network generator process of cartoon transformation. Its value indicates how much the generator G image looks as a cartoon image. This statistic shows the loss of opponents.

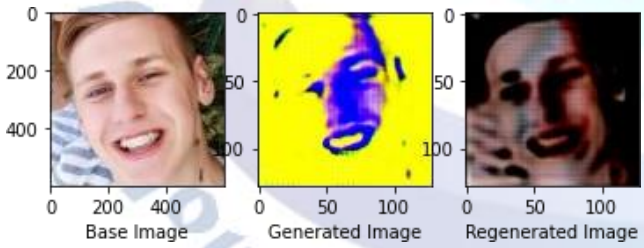

Figure 10: Original photo to generated image

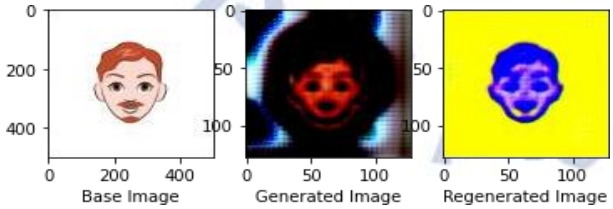

Figure 11: Original cartoon imageregenerated image after loss

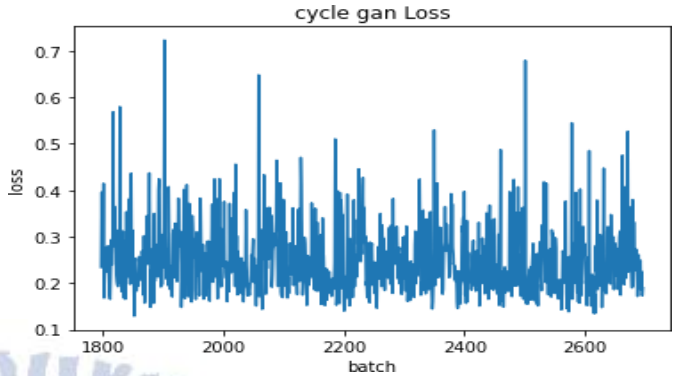

Figure 12**: Graph of cycle gan loss**

The Loss value generated is 0.2826 Considered for five and ten batch sizes.

6. CONCLUSION

We can conclude that the loss experienced during the Cartooninzing of real-time pictures using convolutional UNet architecture is greater than the loss experienced during the Cartooninzing of real-time images using convolutional Resnet design. As a result, we suggest the ResNet as a suitable technique for producing cartoon pictures inside the GAN framework.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] flying face picture memorabilia. 2013 International Computer Vision Conference (ICCV).

[2] Khosla Aditya, Das Sarma Atish and Hamid Raffay. What's popular with an image? Seoul, Korea, April 2014, International World Wide Web Conference (WWW).

[3] C Zitnick Lawrence and Piotr dollar. Edge boxes: Object suggestions are located from the edges. Pages 391-405 of Computer Vision-ECCV 2014. Friday, 2014.

[4] Barnes, C., Shechtman, E., Finkelstein, A., Dan, B.G.: Patchmatch: a structural image editing random correspondence method. ACM Trans. Figure 28(3),1–11. (2009)

[5] D., Fergus, R.: Predicting depth, normal surface areas and semantine labels using a shared convolutionary multi-scale architecture. In: IEEE International Computer Vision Conference. pp. 2650-2658. (2015)

[6] Freeman, W.T., Jones, T.R., Pasztor, E.C.: Super-resolution, Example-based. Graphics & Applications IEEE Computers 22(2), 56–65 (2002)

[7] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing

Systems - Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2672–2680.

[8] Hays, J., Efros, A.A.: million photos completing the scene. Joint ACM 51, 87–94. (2007)

[9] Iizuka, S., Simo-Serra, E., Ishikawa, H.: Let there be colour!: end-to-end learning of global and local picture priors for simulated-taneous image colouring automatically. ACM Trans. Chart 35, 110:1–110:11. Trans (2016)

[10] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Ephro, A.A.: Encoding context: inpainting learning feature. CVPR pages 2536–2544. (2016)

[11] Radford, A., Metz, L., Chintala, S.: Deep generative networks with deeply convolutionary representational learning. CoRR / 1511.06434 abs (2015)

[12] Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutionary networks of deep belief for scalable and uncontrolled learning of hierarchical representations. In: ICML Machine Learning Conference, Montreal, Quebec, Canada, pp. 609-616, 2009. (2009)