International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 01 | Jan 2020 www.irjet.net p-ISSN: 2395-0072

Transformation of Realistic Images and Videos into Cartoon Images and Video using GAN

Akanksha Apte[1], Ashwathy Unnikrishnan[2], Navjeevan Bomble[3], Prof. Sachin Gavhane[4]

[1,2,3]Dept. Information Technology, Atharva College of Engineering, Mumbai.
[4]Dept. Information Technology, Atharva College of Engineering, Mumbai.

Abstract - Aim of the project is to put forward a solution for transforming snapshots or videos of real-world into animated photos (Cartoon Images) or Video. The earlier method of transformation requires complicated computer graphics and skills. The idea of the paper is based on designated snapshots and videos which are converted to an art form such as painting. Amongst all the techniques usable, the application of a Generative Adversarial Network (GAN) called Cartoon GAN will be used for the styling real-world images that use 2 loss functions namely, content loss and adversarial loss for getting a sharp and clear image. With the help of GAN, it is possible to convert video as well to its cartoonized version and the development of the project shows that our Proposed Idea provides high quality cartooned images and videos.

Key Words: Entertainment, Image Processing, Animation, Generative Adversarial Network.

1. INTRODUCTION

Cartoons are commonly used in various kinds of applications. As we know cartoons are artistically made it requires elegant and fine human artistic skills. While portraying Cartoons in humongous numbers for any animated movies it gets time-consuming for the artist as they need to define the sketch of the cartoon properly to get a good result. We all know that animation plays an important role in the world of cinema, so to overcome the problem faced by the artist we have created a program with the help of GAN which not only converts images but also converts video into an animation.

A couple of years ago, the styling of images consists of a particular domain named "non-photorealistic rendering". The Traditional algorithm was developed on the base of the domain for the styling of images and they were successful in styling any images by adding designs, texture, effects, etc. With the help of the algorithm, many software was developed to convert real images(snapshot) into cartoon images some of the methods failed while some of the methods gave results but didn't satisfy all the requirements.

Moreover, cartoon images are byzantine compared to real-life images.

To satisfy all the requirements of converting real images i.e. snapshot into cartoon image we have taken the help of Cartoon GAN [1] which is one of the applications of Generative Adversarial Network (GAN). This program will ease human work. In the above image, a real image is converted into an animated image with the help of the application developed known as "Cartoon GAN" in less duration of time. Not only of the transformation of images but we have also successfully converted video clips into animation with the help of CV2 which is a library in python. In this manner time is saved, fine work is generated within an ample amount of time, which will give a great opportunity for animation industries to make as much as movie or animation clips. This method gives a more accurate result of converting images plus it converts video into an animation clip compared to the previous methods.

2. LITERATURE REVIEW

2.1 An introduction to image synthesis with GAN.

A couple of years back, there had been tremendous growth in the research of GAN (Generative Adversarial Network) [9]. GAN was put forward in the year 2014 where it was introduced in various applications such as deep learning, natural language processing (NLP). From this paper, we explored the different methods of image synthesis such as direct method, Hierarchal method and iterative method [3].

They spoke about two methods of image synthesis which are "text-to-image conversion" and "image-to-image translation". In text-to-image conversion, current methods worked well on a pre-defined dataset where each image contains one object such as Caltech-UCSD Birds [6] and Oxford102 [7], but the performance on complex datasets such as MSCOCO [14] is much poor. While some of the models were successful in producing realistic images of rooms in LSUN [16] because the rooms didn't contain any living things. So, they got a successful image of the room because living things are more complicated to convert then static objects. This was the limitation in this model and it was necessary to learn different concepts of the object.

To improve the performance of GAN and enhance output in the task they trained different models that would generate a single object and train another model which would learn to combine various objects according to text descriptions, and that CapsNet [10].

Now coming to image-image translation, they discussed some general models from supervised to unsupervised settings which are pixel-wise loss [13], cyclic loss [14] and self-distance loss [12]. Aside from this they also proposed some image – image translation model for face editing, video prediction and, image super-resolution. We can say that image-image translation was a catchy application of GAN which had a great scope for mobile application. Although during the research unsupervised method were seen to be more popular compared to the supervised method

2.2 Auto-painter cartoon image generation from sketch using by using cGAN

The authors studied various problems faced by sketch artist while sketching various black and white cartoon drawings, for consideration coloring of various sketches, mixing of a different color to get a unique shade for a particular sketch. According to their research, some difficulties were faced by the artist to get a unique or desired color they need after mixing two or more colors. So to overcome this problem an application was introduced which was known as "sketch-to-image synthesis while using conditional generative adversarial networks" (cGAN)[3]. Later on, they found that the application faced a problem and failed to give the desired output.

To avoid this issue, they invented the Auto-painter model which could automatically generate suitable colors for a sketch. Their application was based on conditional GAN with 'Unet'[2] structure which allowed the output image to have both low-level information of sketch as well as learned high-level color information. They also founded more constraints based on the pix2pix [20] model to obtain finer painting. Here they worked on the auto-painter to alter to color control so that their network could alter the combined result which could satisfy the user by various colors. The result showed that the auto-painter could generate a polished animated image from the two given datasets.

In spite of the guaranteed result, the system suffered from difficulties of adjusting parameters just like other D.Learning models. Also, the combination network structure resulted in less speed for training.

3. METHODOLOGY

GAN i.e. "Generative Adversarial Networks" algorithm is used for the implementation. GAN is a combination of a generative model and a discriminative model. The generative model creates new instances of data that resemble the training data. The discriminator is the model used for testing the data and comparing it with the image from the Generator. Discriminator decides whether the output image is fake or real.[8] The Generator and discriminator both are neural networks and both run in competition with each other in the training phase. The steps are repeated multiple times so that the generator and discriminator get a better result after the repetition of steps. The working can be visualized by the diagram given below:

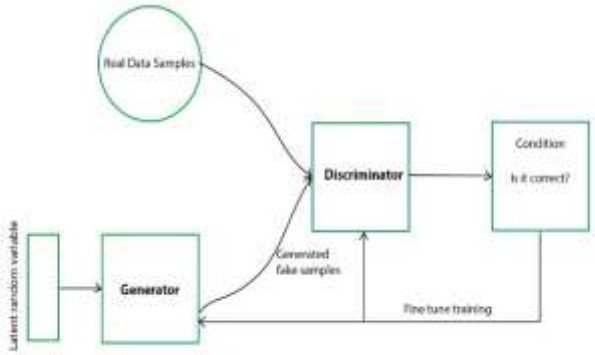

Fig-3.1 Working model of GAN

Now coming to the project, we have used for about 3000 images for training the model from Flickr (Flickr is an image hosting service and video hosting service).

Compare to the previous method of cartoon GAN they have training data that contains the real-world photo and the cartoon image but, in our project, we only train the real-world image that is captured with the help of a camera.

Here the training data consist of (snapshot) real-world images which were passed through 200 epochs, trained [1]to produce the cartoon model with the help of generator were it reconstructs the cartoon image with the help of the adversarial function(that comes other loss function) and the images are trained again under content loss (that comes under loss function).

The final output is generated with the convolution block. Later the discriminator checks if the image generated from the generator is real or fake. To classify whether an image is fake or real is comparatively a less demanding task as compared to generating an image Using GANs is very efficient since it provides high-quality cartoonized images.

To transform videos into animated or Cartoonized videos, we have additionally used the cv2(Computer vision application) which is a library in python the video is divided into frames depending on the specified time period

fps=0.5s

One can change the time period and No. of frames will differ. Later on, these frames will be brought together (joined) by os.join to convert into video files (.mp4 or .avi).

Conversion of Video is basically similar to getting an animated video out of the normal one. The video is first divided into frames and saved into an array, then passed through the generator and discriminator with the help of OpenCV to get the animated(cartoonized video).

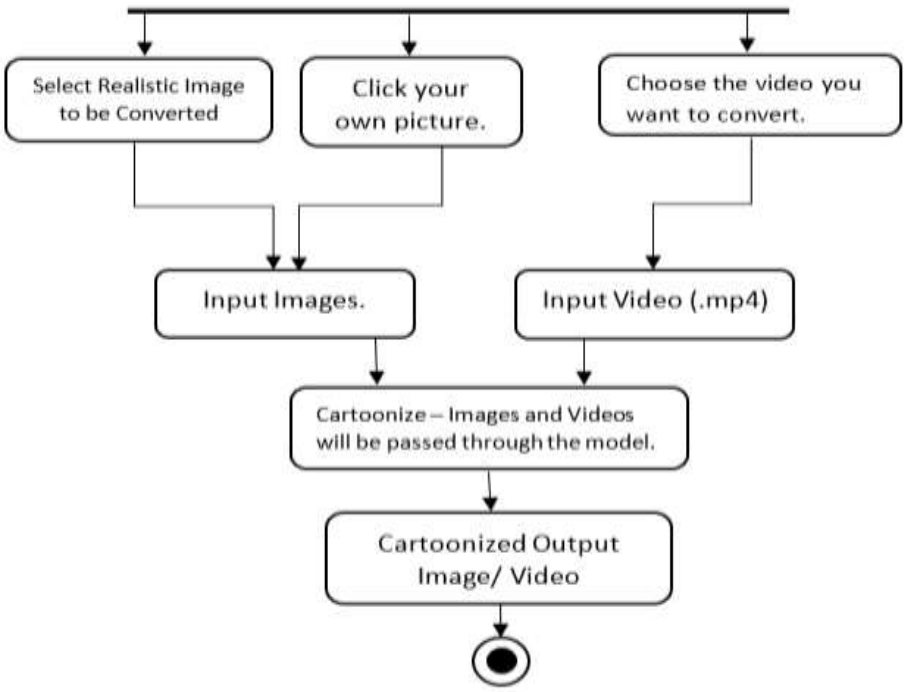

Fig 3.2 Activity Diagram

The above activity diagram shows the working of the project. Where it gives an overall brief idea on how the images and video taken from the camera are converted into cartoon style (Animation).

4. RESULTS

In the below results, the first image consists of the real world which is taken from a video clip and later after using Cartoon GAN with the help of OpenCV we get the cartooned image of the video as you can see below:-

Fig 4.1 Real and Cartoonized Image.

5. CONCLUSION

In this paper with the help of CartoonGAN, where GAN stands for Generative Adversarial Network is used to transform images (snapshots) to the finest cartooned image(animated image). With the help of the loss function and its two types named as Adversarial loss and Content Loss, we got a flexible as well as a clear edge defined images. Also with the help of CV2 which is known as Computer vision, we have transformed video into animation(cartoonized video)

6. FUTURE SCOPE

The project showed that image was successfully converted into a cartoon-style image with help of Cartoon GAN as mentioned in [1]also the video clips were transformed into an animation clip with the help of the python library called cv2.

In the future work, we would like to focus more on generating a portrait defined HD image even though we used the loss function but still failed to the result . We also plan on focusing more on the video conversion so we get HD or a 4k quality video which will be more beneficial

ACKNOWLEDGEMENT

We have tried our best to present Paper on the "Transformation of realistic images and videos into Cartoon images and video by Using GAN" as clearly as possible. We are also thankful to our Guide Prof. Sachin Gavhane for providing the technical guidance and suggestions regarding the completion of this work. It's our duty to acknowledge their constant encouragement, support and, guidance throughout the development of the project and its timely completion.

We are also thankful to Dr. S. P. Kallurkar (Principal), Prof. Deepali Maste (HOD, Information Technology Engineering), without their support and advice our project would not have shaped up as it has.

REFERENCES

[1] Y. Chen, Y.-K. Lai, Y.-J. Liu, "CartoonGAN: Generative Adversarial Network for photo cartoonization", International Conference on Image Processing, 2018

[2] Y. Chen, Y.-K. Lai, Y.-J. Liu, "Transforming photos to comics using convolutional neural networks", International Conference on Image Processing, 2017

[3] Zengchang Qin, Zhenbo Luo, Hua Wang, " Auto-painter: Cartoon Image Generation from Sketch by Using Conditional Generative Adversarial Networks", International Conference on Image Processing, 2017

[4] J. Bruna, P. Sprechmann, and Y. LeCun., "Super-resolution with deep convolutional sufficient statistics" In International Conference on Learning Representations (ICLR), 2016

[5] K. Beaulieu and D. Dalisay, "Machine Learning Mastery", Machine Learning Mastery, 2019. [Online]. Available: https://machinelearningmastery.com/. [Accessed: 24- Nov- 2019].

[6] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The caltech-ucsd birds-200-2011 dataset," 2011.

[7] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing, Dec 2008.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft coco: Common objects ´ in context," in European conference on computer vision. Springer, 2014, pp. 740–755.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, 2014, pp. 2672–2680.

[10] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," arXiv preprint arXiv: 1710.09829v2, 2017.

[11] Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. CoRR abs/1611.07004 (2016)

[12] S. Benaim and L. Wolf, "One-sided unsupervised domain mapping," arXiv preprint arXiv:1706.00826, 2017.

[13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," arXiv preprint arXiv:1611.07004, 2016

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired imagetoimage translation using cycle-consistent adversarial networks," arXiv preprint arXiv:1703.10593, 2017.