International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.034 ISSN: 2581-4621

Effective Mode of Learning Cartonization: White-box Cartoon Representations

Shivendu Anand[1]
[1]*UG - Information Science Engineering, RV College of Engineering, Bangalore, Karnataka*

ABSTRACT
Cartoon, a renowned art form has different forms of application in various scenarios. In essence, different cartoon styles and use cases, demand prior knowledge to develop a usable algorithm or they have certain assumptions that needs to developed which cater to a particular task. The present paper's objective is to determine an impactful approach towards image cartoonization, by developing an efficient and effective mode of such learning. The author proposes to identify three white-box representations from images individually, namely, the surface representation that covers a smooth surface of cartoon images, the structure representation that observes the sparse color-blocks and the texture representation which observes high- frequency texture, contours, details in cartoon images. The objectives of the proposed method are separately based on each extracted representations, that applies a Generative Adversarial Network (GAN) framework to understand the extracted representations and cartoonize images, which further enables the framework to be under control and adjustable at the convenience of the author. The present approach aspires to benefit the requirements of an artist that deals in several cases and has various style requirements. The author has conducted analytical and doctrinal studies, coupled with bibliometric research, that covers both quantitative analysis and qualitative comparisons, accompanied with the empirical study that have been present in several research papers, related to the subject have been thoroughly analyzed by the researcher to add something to existing knowledge on White box Cartoon Representations. Lastly, it is observed that the ablation study carried out shall showcase an influence of each component in the framework suggested by the author.
Keywords-Image Cartoonization, White-box cartoon Representations, Generative Adversarial Network (GAN).

1. Introduction

In the modern age, cartoon animation workflow is crucial as it allows artists to allow numerous sources that enables them to create content with the upcoming trends. The process of image cartoonization involves the usage of famous products that have been captured using real -world photography into that of cartoon scene materials that are usable for content creation. Image Cartoonization involves the usage of several cartoon styles and depends on the usage of several representations that depend on a particular task or is carried out by a prior assumption undertaken to develop usable algorithms.

The cartoon workflows have different requirements, depending on a case to case basis, wherein one focusses on global palette themes, where the sharpness of lines becomes a trivial issue, whereas in other workflows, clean and sparse colours are preferred in a piece of art over that of the theme of the workflow. Such different variants poses severe issues to black-box models, (1) as diverse demands of artists are to be catered in such issues, and the change of training dataset that is a simple solution, is generally not opted.

Cartoon Generative Neural Networks (GAN) framework,(2) is designed for image cartoonization, wherein a novel edge loss is developed that achieves outstanding results in certain cases, however, using a black-box model for the training data directly, shall decrease the generality and stylization quality, leading to bad cases. The author through this paper proposes three cartoon representations based on several observations of cartoon painting behaviour; firstly, the surface representation, secondly the structure representation, and lastly, texture representation.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.034 ISSN: 2581-4621

2. Similar Work

2.1. Image Smoothing and Surface Extraction

Image smoothing, (3) is an extensively researched topic wherein early methods are mainly filtering based, and optimization-based methods became recognised at a later stage. Farbman, (4) utilized weighted least square to constrain the edge-preserving operator and Min,(5) solved global image smoothing by minimizing a quadratic energy function, In this paper, the author strives to adapt to a differentiable guided filter, that will enable the user to extract smooth, cartoon-like surface from images, thus becoming an effective structure-level composition and providing a smooth surface that assists artists in making creative art cartoons.

2.2. Generative Adversarial Networks

Generative Adversarial Network, (6) is a phenomenal generative model that generates data with a similar distribution system containing input data, by solving a minimum-maximum problem that occurs between a discriminator network and a generator network. This particular tool is extremely significant in the process of image synthesis, as it forces the generated as similar to that of real images. GAN has recognised in several conditional image generation tasks, such as image inpainting, image colorization, style transfer and image cartoonization, The author proposes to adopt an adversarial training architecture, with the usage of two discriminators that enforces the generator network to synthesize images, that would use the same distribution network selected required as the domain in target.

2.3. Super pixel and structure Extraction

Super-Pixel Segmentation groups,(7) are connected pixels in an image contain similar colour or grey level in their group. Certain popular superpixel algorithms are based on graph, that treats pixels as nodes and any connection between pixels are treated as edges in a graph. The author observes that the gradient ascent based algorithms begins the process of formation of image with rough clusters and optimises the present cultures by grouping them with gradient ascent until convergence. In this present paper, the author has used the felzenszwalb algorithm,(8) that will be used to develop a cartoon-oriented segmentation method, which aids in capturing a learnable structure representation. Such a representation shall be an advanced step for deep models that require seizure of content that is available throughout the world, and enables to produce usable results that can be practically implemented for achieving celluloid style cartoon workloads.

2.4. Non-photorealistic Rendering

Non-photorealistic Rendering (NPR) methods seeks to represent image content that use different artistic styles, like pencil sketching, paints and usage of water colour. Image cartoonization is studied by adopting a filtering based method that enables an end-to-end neural network, and covers the use of portraits, photos and videos. The optimisation of a style loss and a content loss was propounded, that would generate stylize images, which was also recognised as style-content image pair.(9) However, it was later propounded by several works that proposed different methods to style images. The range of application of NPR is tremendous in the process of image abstraction. (10)The use of semantic edges while filtering the details present in the image, enables to present abstracted visual information that are commonly observed in cartoon related applications. The author in the present paper strives to adopt a different approach from style transfer methods which uses a single image or image abstraction methods that take into account of simpler content images and strives to learn the cartoon data distribution from a set of cartoon images, thus allowing the model to capture high quality cartoonised images.

2.5. Image-to-Image Translation

Image-to-Image Translation(11) strives to address the issue of translating images from a source domain to another target domain. The applications containing such translation include image quality enhancement,(12) stylizing photos into paints, cartoon images and sketches. At present, it was

DOI:10.46647/ijetms.2022.v06i05.034 ISSN: 2581-4621

observed that bi-directional models are often introduced for interdomain translation. Through this paper the author adopts an unpaired image-to-image translation framework for image cartoonization.

3. Approach Proposed by the Author

The author describes the process of image cartoonization framework wherein the images are decomposed to the extent of the three categories previously states, which are surface representation, the structure representation, and the texture representations. These modules are independent that are introduced to extract representation that use the GAN framework. Furthermore, the Pre-trained VGG network is required to extract high-level features that imposes a spatial constrain, that would showcase global contents between extracted structure representations and outputs, and also between input photos and outputs.

3.1. Outcome of Surface Representation

It was rightly observed that the surface representation resembles the cartoon painting style where the artists draws an estimated draft with coarse brushes, having smooth surfaces, which are similar to that of cartoon images. Furthermore, in order to smooth images keep up with the global semantic structure, a differentiable guided filter is generally preferred for edge preserving filtering.

3.2. Outcome of Structure representation

The outcome of the Structure representation observed is that of a flattened global content, accompanied with sparse colour blocks, that clears pre-established boundaries in celluloid style cartoon workflow. The author undertakes the usage of the felzenszwalb algorithm, that assists in segmenting images into separate regions. Furthermore, such super pixel algorithms are only considered, that have similarity of pixels, which also ignores semantic data. The author also introduces selective search, (13) that would assist in merging segmented regions and extracting a sparse segmentation map. It is observed that the standard super pixel algorithms colour of each segmented region with an average of the pixel value is recorded with utmost care and diligence. Lastly, it is perceived that by analysing the processed dataset, the author inferred that this lowers global contrast as well as darkens images, which causes a hazing effect on the final results.

3.3. Outcome of Textural Representation

The author records the outcome of textural representation, wherein the high-frequency features of cartoon images are recognised as significant learning objectives, however the luminance and colour data of the piece of art makes it easy to distinguish between cartoon images and real-world photos. Thus, the author proposes that a random colour shift algorithm should be adopted that would extract single-channel texture representation from colour images, which not only retains high-frequency textures, but also decreases the influence of colour and luminance.

4. Case Study

4.1. Illustration of Controllability

With the adoption of the proposed doctrinal method suggested by the author, the style of cartoonized results was viewed to be adjusted by choosing to turn the weight of each representation in the loss function. Furthermore, it was recognised that by increasing the weight of texture representation, it adds more details in the illustrated images, which are filled with rich details such as grassland and stones, which are preserved. The author believes that the regulation of dataset distributions, enhances high-frequency details stored in texture representation.
The illustration of our method being adopted as per the cartoon images (right) and compared with the real world images (left), in the same scene has been showcased for reference.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.034 ISSN: 2581-4621

Fig 1: Comparison of same scene as a real image with that of our proposed method

Fig 2: Comparison of same scene as a real image with that of our proposed method

Additionally, it was observed that smoother textures and fewer details are generated with a higher weight of surface representation, The reason behind such a pattern is that the guided filtering smooths training samples and reduces densely textured patterns. Thus, in order to get more abstract and sparse features, the author increases the weight of structure representation, because the selective search algorithm reduces the training data into structure representations. Hence, unlike other black-box models, the white-box method suggested by the author is easily adjustable and controllable.

4.2. Qualitative Comparison

The comparisons between the method suggested by the author previous methods are shown below-

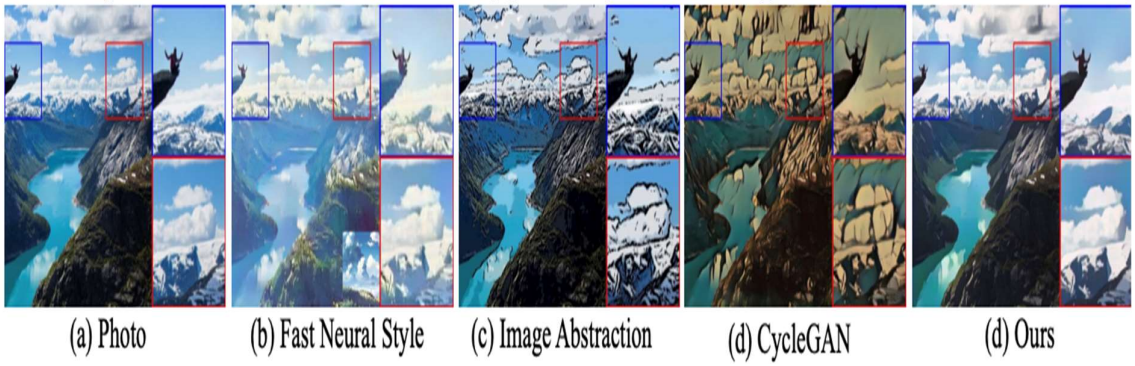

Fig 3: Comparison between various methods to that our proposed method

The author recognises several improvements with the adoption of the white-box framework, which is that of generate clean contours. It is pertinent to that that the image abstraction causes noisy and messy contours, wherein as observed in the other previous methods, there is a lack of clear borders, as compared to clear borders observed in the method suggested by the author. Furthermore, it is rightly pointed out that Cartoon representations aid in maintaining harmonious colouring.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.034 ISSN: 2581-4621

The CycleGAN approach generates darkened images and sass neural style causes colour to become oversmoothed, which distorts colours like human faces and ships. The method suggested by the author, prevents improper colour modifications that enables artists to depict their art in a clearer format. Lastly, the method effectively reduces insignificant information while preserving fine details, but all the other methods either causes over-smoothed features or distortions.

4.3. Quantitative Evaluation

Frechet Inception Distance (FID), (14) is abundantly used to quantitatively evaluate the quality of synthesized images. The Pre-trained Inception-V3 model, (15) is used to extract highlevel features of images and calculate the distance between two image distributions. The author uses FID model to evaluate the performance of other methods in comparison to the method suggested by the author. As CartoonGAN models have not been tried on human face data yet, for fair comparisons, the author uses FID approach to be calculated on a scenery dataset. It was observed the method suggested by the author generates images with the smallest FID to cartoon image distribution, that establishes the generation of results, which are most similar to cartoon images. Lastly, the output of the method also had the smallest FID to real world photo distribution, which indicates that the suggested method loyally preserves image content information.

5. Conclusion

In this paper, it was witnessed that the images being decomposed into three cartoon representations; the surface representation, the structure representation, and the texture representation and corresponding image processing modules are used to extract three representations for network training, and output styles could be controlled by adjusting the weight of each representation in the loss function. The authors had proposed a white-box controllable image cartoonization framework, which is based on GAN, that shall generate high-quality cartoonized images from real-world photos. A GAN-based image cartoonization framework shall be optimized with the guide of extracted representations. The use of the method suggested by the author enables the users to adjust the style of model output by balancing the weight of each representation. Lastly, it was inferred that the suggested method outperforms existing methods in qualitative comparison and quantitative comparison.

References

1. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision, pages 694–711. Springer, 2016.
2. Yang Chen, Yu-Kun Lai, and Yong-Jin Liu. Cartoongan: Generative adversarial networks for photo cartoonization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 9465–9474, 2018.
3. Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In European Conference on Computer Vision, pages 1–14. Springer, 2010
4. Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation.
5. Dongbo Min, Sunghwan Choi, Jiangbo Lu, Bumsub Ham, Kwanghoon Sohn, and Minh N Do. Fast global image smoothing based on weighted least squares. IEEE Transactions on Image Processing, 23(12):5638–5653, 2014.
6. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems, pages 2672–2680, 2014.
7. Greg Mori. Guiding model search using segmentation. In Proceedings of IEEE International Conference on Computer Vision, volume 2, pages 1417–1423. IEEE, 2005.

DOI:10.46647/ijetms.2022.v06i05.034 ISSN: 2581-4621

8. Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. International Journal of Computer Vision, 59(2):167–181, 2004.
9. Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576, 2015
10. Jan Eric Kyprianidis and Jurgen D ¨ ollner. Image abstraction ¨ by structure adaptive filtering. In TPCG, pages 51–58, 2008.
11. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei 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
12. Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: weakly supervised photo enhancer for digital cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 691–700, 2018.
13. Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. International Journal of Computer Vision, 104(2):154–171, 2013
14. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Advances in Neural Information Processing Systems, pages 6626–6637, 2017
15. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2818–2826, 2016.