GANmera: Reproducing Aesthetically Pleasing Photographs using Deep Adversarial Networks

Nelson Chong Lai-Kuan Wong John See
Faculty of Computing and Informatics, Multimedia University
Persiaran Multimedia, Cyberjaya, Selangor, Malaysia.
nelsonchong9323@gmail.com {lkwong,johnsee}@mmu.edu.my

Abstract

Generative adversarial networks (GANs) have become increasingly popular in recent years owing to its ability to synthesize and transfer. The image enhancement task can also be modeled as an image-to-image translation problem. In this paper, we propose GANmera, a deep adversarial network which is capable of performing aesthetically-driven enhancement of photographs. The network adopts a 2-way GAN architecture and is semi-supervised with aesthetic-based binary labels (good and bad). The network is trained with unpaired image sets, hence eliminating the need for strongly supervised before-after pairs. Using CycleGAN as the base architecture, several fine-grained modifications are made to the loss functions, activation functions and resizing schemes, to achieve improved stability in the generator. Two training strategies are devised to produce results with varying aesthetic output. Quantitative evaluation on the recent benchmark MIT-Adobe-5K dataset demonstrate the capability of our method in achieving state-of-the-art PSNR results. We also show qualitatively that the proposed approach produces aesthetically-pleasing images. This work is a shortlisted submission to the CVPR 2019 NTIRE Image Enhancement Challenge.

1. Introduction

As technology continues to develop, the use of digital photography has been increasingly prevalent in the modern society. This increases the need to enhance the aesthetic quality in images, be it for commercial or personal uses. Image editing software, such as Adobe Photoshop, Affinity Photos and even the open source GIMP, have been introduced over the years as image editing solutions for professionals in the field. However, this editing process requires substantial amount of skill, patience and time. On the contrary, users unable or unwilling to spend their time and money on these software can use pre-sets and filters offered by apps like Instagram and Prisma to stylize their photos. However, these filters mostly utilize global-based image enhancement techniques and often fail to produce results to the satisfaction of the users.

Aesthetics image enhancement is often a challenging task, largely due to the empirical and perceptual process of photo editing as well as the need for avoiding artefacts and preserving naturalness of an enhanced photograph. With the recent leap frog advancement of the deep learning techniques, particularly General Adversarial Networks (GAN), several researchers [7] [5] [4] [8] have explored the use of GAN-based networks for automatic image enhancement. These approaches produce some impressive results but are not without limitation. One common problem with these methods is noise amplification. In addition, [7]'s method are prone to color deviation and over-contrast while [4] noted the tendency of halo effect in their results. On the other hand, EnhanceGAN [5] can produce impressive and dramatically enhanced results, but sometimes at the expense of over-saturation and naturalness of the images.

In this paper, instead of proposing a new network architecture, we focus on fine-grained adjustment of a state-of-the-art GAN-based network to overcome the aforementioned limitations. We model the image enhancement problem as an unpaired image-to-image translation problem. Specifically, we adopt the 2-way GAN architecture based on CycleGAN [29] with modifications made to the loss function, activation functions and resizing schemes. We also explore dual-stages training strategy, where more than one training stages and training datasets are employed to train the network. Quantitative and qualitative experiments performed on our model illustrates the capability of our proposed method in producing high-quality enhanced photographs with natural colour rendition.

2. Related Work

Pioneering methods of image enhancement focus either on contrast enhancement [18] [23], color correction [20]

[14] or sharpening [17] and are non-aesthetics driven. Due to the lack of consideration for the content of the image, these methods oftentimes fail to improve image aesthetics significantly. Notably, an important aesthetics guideline is making a photo subject dominant such that viewers can be directed to what the photographers want them to see. Inspired by this aesthetics characteristics, several researchers explore various techniques [25] [13] [16] to perform region-based low-level image enhancement that attempt to alter the saliency of a photograph by applying non-linear low-level enhancement, including contrast, color and sharpness enhancement to different objects in the image with the aim to make the main photo subjects more dominant. The drawback of these approaches is the need for prior contextual information such as object probability map or scene descriptors, thus reducing the robustness and scalability of these techniques.

Another approach to enhancing a photograph is to transfer the color style from photographs of professional photographers. Earlier methods [19] achieve color correction by choosing an appropriate source image and apply its color characteristic to the input image. Following that, [12] performed exemplar-based color transfer by automatically retrieving matching exemplars from collection of photographs onto the given input image. This approach is subject to erroneous results if exemplar matching failed. [28] casts exemplar-based photo adjustment as a regression problem, and use a deep neural network (DNN) to represent the highly nonlinear and spatially varying color mapping between input and enhanced images. Their approach relies on both scene parsing and object detection to build semantic label maps. Notably, mislabeling in the semantic map can propagate into contextual features and adversely affect photo adjustment.

More recently, with the introduction of Convolutional Neural Network (CNN), particularly Generative Adversarial Network (GAN) that has the innate generative ability, several GAN-based image enhancement methods emerged. Earlier GAN methods require paired images to be used for supervised learning of the enhancement and directly alter individual pixels in an image. [7] presents a novel photo enhancement approach that learns a mapping between photos from mobile devices and a DSLR camera with the aim to produce DSLR quality photos. They introduced DPED, a new large-scale dataset of 6000 photos taken synchronously by a DSLR camera and 3 low-end cameras of smartphones. The authors extended their work to WESPE [8], a generic method capable of enhancing source images into DSLR-quality in a weakly-supervised manner, compared to the strongly-supervised DPED [7]. [4] DPE capitalized on the strength of a 2-way GAN approach, whereby paired images are not required for training. On the other hand, instead of having the network directly modifying the pixels, EnhanceGAN [5] uses a parameterized approach. They trained their model in a weakly-supervised manner using only images with binary good and poor labels, without the need of paired-image dataset. While these approaches generate some impressive results, several limitations still exists including amplification of noise, color deviation and halo effects.

Considering the importance of efficiency of the enhancements methods for practical usage, the PIRM challenge on perceptual image enhancement on smartphones [10] aimed to benchmark resource-efficient architectures targeted at high perceptual results and deployment on mobile devices. Notably, winning methods of the image enhancement track of the challenge were able to significantly improve the runtime and PSNR scores of DPED. Mt.Phoenix, the winner of the challenge proposed a U-net style architecture and augmented it with global features. Their method produced the highest perceptual score coupled with the fastest runtime on both CPU and GPU. EdS proposed some modifications to the convolutional layers of the DPED ResNet architecture for faster training and managed to achieved the second best perceptual score. On the other hand, MENet that achieved the highest PSNR and MS-SSIM scores in the challenge proposed a θ-inception network that has a θ-inception block, where the image is processed in parallel by convolutional an deconvolutional layers to achieve high efficiency.

3. Proposed Approach

The base architecture of the proposed method is inspired by CycleGAN [29], which was originally designed for performing advanced image-to-image translation without paired dataset. CycleGANs architecture was chosen due to its 2-way structure and its cyclic consistency loss. In a 2-way GAN structure, an image is first forward-mapped from its original domain to the target domain, then backward-mapped back to the original domain. The cyclic consistency loss enforces the rule that the output image and the original image must share some common features, and must be "structurally" similar. This allows the network to learn and map features more meaningfully and effectively than traditional 1-way structures. Several fine-grained modifications were made to the original CycleGAN method, including the loss function, activation functions and resizing method in order to produce aesthetically pleasing enhanced photographs. The proposed architecture is similar to WESPE [8] as they do not require paired datasets and have a loss function dedicated to measuring the difference between the original input and the reconstructed input. However, WESPE features 5 fully convolutional networks: 2 generators, 2 discriminators and 1 pretrained VGG-19 [21] for measuring content loss. Although their method also feature 2 discriminators, their discriminators are tasked with measuring different losses, respectively: color loss and texture loss.

Based on the CycleGAN architecture, we formulate the image enhancement problem as an unpaired image-to-image translation problem. Let X and Y be the domain of "bad" and "good" images respectively. Given training samples comprising of a set of bad images, $\{x_i\}_N^{i=1}$ where $x_i \in X$ and a set of good images, $\{y_i\}_M^{i=1}$ where $y_i \in Y$, we learn two color mapping functions, G and F between the two domains,

$$G : X \longrightarrow Y \tag{1}$$

$$F : Y \longrightarrow X \tag{2}$$

The color mappings functions, G and F are associated with the adversarial discriminators D_Y and D_X respectively. D_X encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_Y for F. To regularize the color mapping, we enforce two cyclic consistencies, the forward cyclic consistency,

$$x'' = F(G(x)) \approx x \tag{3}$$

the backward cyclic consistency,

$$y'' = G(F(y)) \approx y \tag{4}$$

in order to preserve the structural similarity, where $x \in X$ and where $y \in Y$. With these learnt mapping functions, our proposed method can transform a given input image into an output image, with color properties matching that of the set of good images, Y.

The objective function consists of two types of terms; (1) adversarial losses, and (2) cyclic consistency losses. For the adversarial losses, we use the default mean squared error (MSE) used in CycleGAN. For the cyclic consistency losses, we use the root mean squared error (RMSE) instead of the traditional $L1$ error used in the original CycleGAN network. The cyclic consistency losses, C, is thus defined as,

$$C = \mathbb{E}x, x''[RMSE(x, x'')] + \mathbb{E}_{y,y''}[RMSE(y, y'')] \tag{5}$$

where

$$RMSE(x, y) = \sqrt{\frac{1}{mn} \sum_{i=1} \sum_{j=1} (x(i, j) - y(i, j))^2}, \tag{6}$$

and (i,j) represents the pixel indices. As RMSE takes the square root of the average squared errors, it penalizes large errors more, which is useful for our purpose as the output image is desired to be as structurally similar to the original image as possible.

3.1. Generators

The proposed network contains two generators, i.e. good-to-bad and bad-to-good, both seeking to learn the color mapping functions, F and G respectively. The bad-to-good generator plays an important role as it is used as the final image enhancer. The architecture of the generators are adapted from the generator networks of CycleGAN [29], with several fine-grained modifications made to the loss functions, activation functions and resizing schemes. The architecture can be separated into three parts: down-sampling, residual learning, and up-sampling as shown in Fig. 1. The size of the input image is set to 128×128 while the number of features in the initial layer is set to 64. Every convolutional layer in the generator is activated with a LeakyRelu activation [26] instead of the default Relu function used in CycleGAN, except for the final layer, which is activated with a *tanh* activation since the input is normalized to between -1 and 1. Every convolutional layer is also followed by an instance normalization [24]. Instance normalization was selected over standard batch normalization as batch normalization is likely to add extra noise to the training process, which could hurt the generated output.

The down-sampling part consist of four convolutional layers. The first convolutional layer consists of 7×7 filters with stride 1, while each of the following layers have 3×3 filters with stride 2. The residual learning part consist of six residual blocks, each consist of two convolutional layers with 3×3 filtering and stride 1. Residual learning is used as it helps with convergence and has been proven to be effective at general image processing tasks. These residual blocks also ensure that our generator only learns the difference between the input image and the label images. Instead of using deconvolutional layers in the up-sampling part, we use resize-convolutional layers [1] to up-sample the previously down-sampled images. The problem with deconvolution is that if the kernel size is not divisible by the stride, it could cause checkerboard artefacts. To avoid these artefacts, we took the resize-convolution approach, which involves resizing the image with nearest-neighbour interpolation before performing a convolution. This kind of approach has been known to be robust against artefacts. The final layer is a convolutional layer containing 7×7 filters with stride 1 followed by instance normalization. This layer is activated by a *tanh* function, with glorot uniform initialization [6] and $L1$ regularization [27]. When weights in a network start either too small or too large, the information would shrink or grow as it passes through each layer until it is too tiny or too large to be useful. The glorot uniform initialization (also known as the Xavier initialization) draws the initial weights based on a good variance for the distribution of the data. The $L1$ regularization is added to avoid over-fitting, producing a model that has a sufficiently feasible subset of input features.

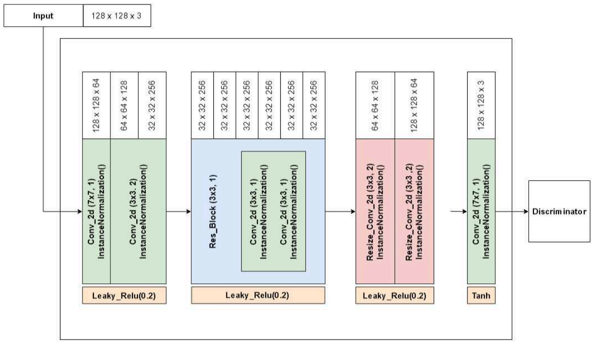

Figure 1. Structure of the Generator network

3.2. Discriminators

For the discriminator networks, we use 60 x 60 PatchGANs [11] as illustrated in Fig. 2. Each discriminator consists of five convolutional layers. Similar to the generators, the size of the input image is set to 128×128 while the number of features in the initial layer is set to 64. The first convolutional layer contains a 4×4 filter with stride 2, while every subsequent layer aside from the final layer contains the same filter-stride setting, and are each followed by instance normalization. Every layer aside from the final layer is activated with a LeakyRelu activation. The final layer is a convolutional layer with 4×4 filter and stride 1 followed by an instance normalization. This layer is set to produce a one-dimensional output in the range of $[0, 1]$ with 0 signifying real and 1 signifying fake. The desired output of the discriminator is to be as close to 0 as possible.

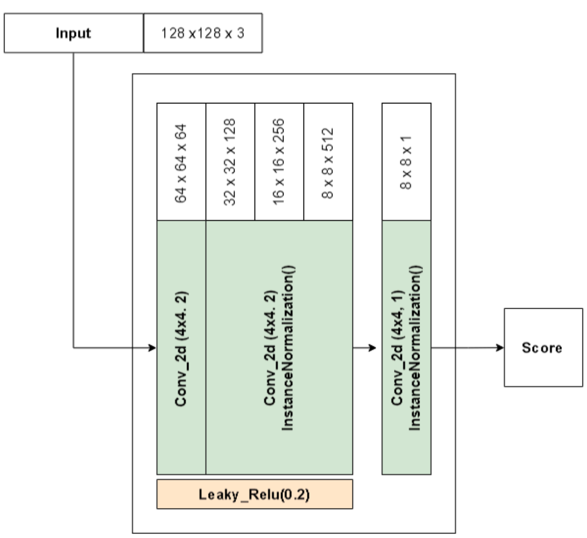

Figure 2. Structure of the Discriminator network

Figure 3. Comparison of our models.

4. Experiments

4.1. Datasets

The main aim of our work is to be able to produce enhanced images with an improved aesthetics appeal. Whilst our participation in the NTIRE challenge warrants the use of its own dataset, we further evaluated our proposed enhancement method on the MIT-Adobe 5K, an existing benchmark dataset with additional consideration for the aesthetic appeal of images.

Aesthetics Datasets. Three aesthetic-centric datasets were chosen to train our aesthetics-driven models: CUHK-PhotoQuality (CUHK-PQ) [22], AVA [15] and MIT-Adobe 5K (MIT-5K) [2]. In these datasets, we focus only on images from outdoor-related categories such as landscape, cityscape and others. CUHK-PQ consists of 4,072 high-quality images and 11,812 low-quality images, with images from 8 categories. We extracted ≈2,600 images from the Architecture and Landscape categories of the CUHK-PQ dataset, with an equal proportion of ≈1,300 images of both high and low quality. The AVA dataset contains 250K images, each labelled with a series of aesthetic ratings from 1 (low) to 10 (high). About 1,300 images with an average rating of greater than or equal to 6 are extracted from the

Figure 4. Another comparison of our models.

Method	PSNR	SSIM
1-stage (hsv)	23.07	0.75
1-stage (lab)	23.43	0.75
1-stage (rgb)	25.54	**0.83**
2-stages (hsv)	23.77 (+0.7)	0.80 (+0.05)
2-stages (lab)	25.54 (+2.11)	0.82 (+0.07)
2-stages (rgb)	**26.08** (+0.54)	0.82 (-0.01)

Table 1. Comparison of models trained with different color space and different training strategies.

Landscape, Seascape, Cityscape, Sky and Water categories to be used as the set of good images. The MIT-5K consists of 5,000 raw images. Since images from this dataset are untouched images, we extracted ≈1,300 images from this dataset to be used as the set of low quality images for training our models. We ensure that they do not overlap with the reserved evaluation samples, which consists of the same 500 images used in DPE [3].

NTIRE Dataset. A new dataset was provided exclusively for the NTIRE 2019 single-image enhancement challenge. This dataset contains real low and high quality paired images (DPED [**?**]) captured with a DSLR Canon camera and an iPhone camera. The dataset consist of 16,000+ low-quality and high-quality images, divided into 3 partitions: training, validation and testing. The testing set consists of 3,057 low-quality iPhone-captured images. All of the images are sized at 100×100, with the majority of the images being a part of a building or a tree. The low-quality and high-quality pairs are not drastically different from each other with minor sharpness and color differences.

4.2. Experiments

We conducted some experiments to investigate the effects of using different colour spaces and the impact of single-stage and dual-stage training on the performance of image enhancement. All models are evaluated on the reserved 500 test images from MIT-5K dataset.

Color space. Experiments were conducted by training our proposed method on good and bad images from the CUHK-PQ dataset on 3 different colour spaces: HSV, LAB and RGB.

Single-stage vs Dual-stages training. For the single-stage training strategy (denoted as *1-stage*), training is performed solely on the CUHK-PQ dataset, with the high-quality set being the target domain (good images) and the low-quality set as the source domain (bad images). For the dual-stage training strategy (*2-stages*), we employ 2 stages of training with two separate datasets The CUHK-PQ images are used as both the good and bad images in the 1-stage setting while in the 2-stages setting, we continue refining the enhancement by training further with the MIT-5K and AVA datasets (being the source and target domains respectively).It is important to highlight that, for both training strategies, all datasets used did not provide paired image sets.

Discussion. Table 1 results show the average PSNR and average SSIM scores for our proposed models based on the combination of the three chosen color spaces, with single-stage and dual-stage training strategies. In general, 2-stages models demonstrate better average PSNR and SSIM scores than their corresponding 1-stage models, except for the 2-stages RGB model which had only an insignificant decrease (-0.01) in the average SSIM score. Interestingly, the RGB models outperform their counterparts in both the 1-stage and 2-stages strategies, both quantitatively and visually. From results in Figure 3, 4 and 5, we can observe that all the 1-stage models provide a boost in color saturation and contrast, but sometimes at the expense of over-saturation. Visual comparison of the 1-stage and 2-stages models shows that the dual training strategy is able to lessen the tendency of over-saturation in 1-stage models to provide less dramatic and more natural results. However, we note that the perceptual preference for saturation can differ significantly among individuals and as such, we believe both models can still be good options for image enhancement.

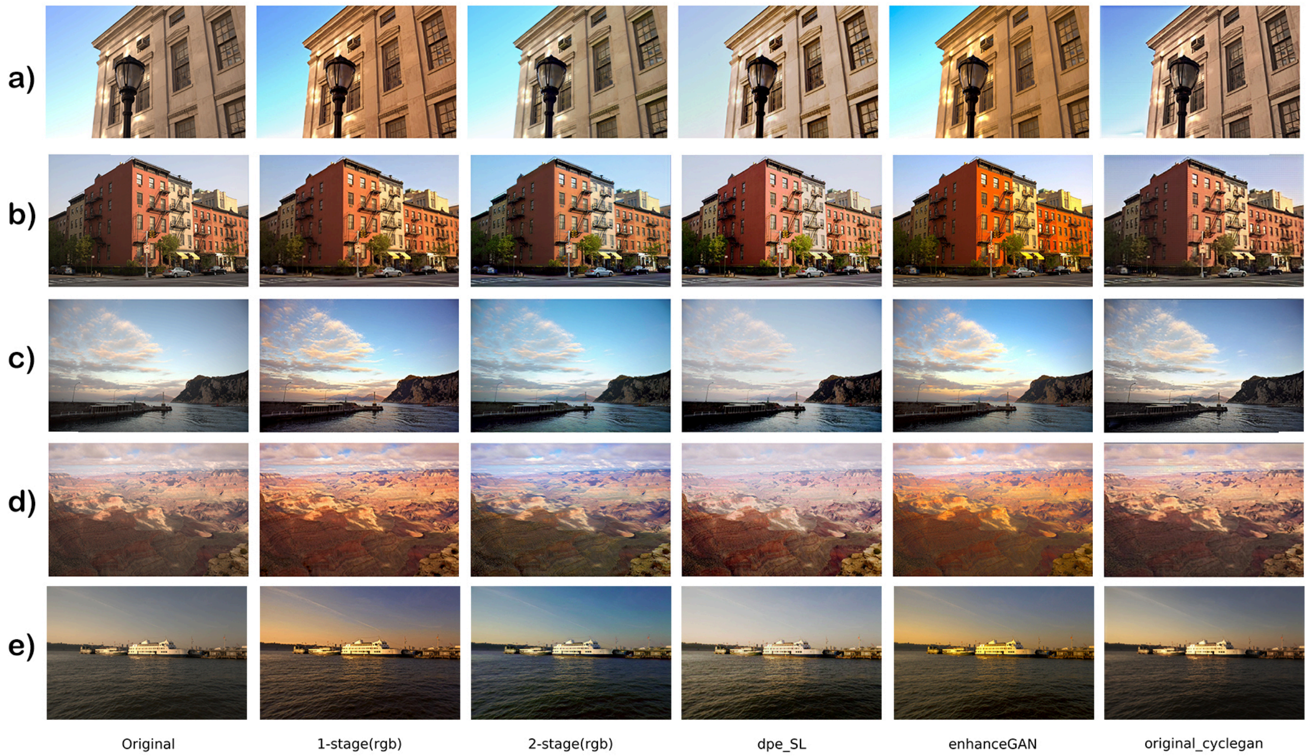

Figure 5. Comparison of our RGB models with CycleGAN [29], EnhanceGAN [5] and DPE [3].

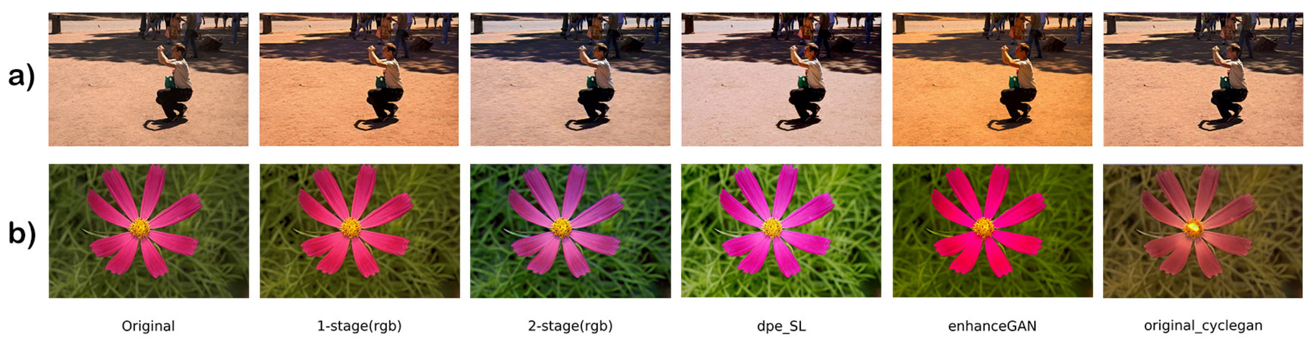

Figure 6. Comparison of our RGB models with CycleGAN [29], EnhanceGAN [5] and DPE [3] on images from categories not trained by our models.

5. Results

5.1. MIT-5K Evaluation

The performance of our models was evaluated against CycleGAN, the baseline architecture which motivated our work, and various state-of-the-art aesthetics-driven image enhancement methods; DPED [7], EnhanceGAN [5] and DPE [3].

Quantitative Evaluation. Table 2 depicts the comparison of the average PSNR and average SSIM scores for our 1-stage and 2-stages RGB models against recently proposed image enhancement models. The results show that both our proposed strategies outperform the CycleGAN approach. This demonstrates that the modifications made to the original CycleGAN (designed for image domain transfer), particularly the loss function, activation functions and resizing scheme, have significantly improve its capability for image enhancement. The superiority of our models in terms of the PSNR score indicate their ability in reducing the problem of noise amplification. The structural similarity of our models against the reference quality level is slightly lower than DPE and DPED but higher than EnhanceGAN

Method	PSNR	SSIM
CycleGAN [29]	21.42	0.62
DPED [7]	21.76	0.87
EnhanceGAN [5]	22.97	0.80
DPE [3]	23.80	**0.90**
Ours - 1-stage (rgb)	25.54	0.83
Ours - 2-stages (rgb)	**26.08**	0.82

Table 2. Comparison of our proposed models against state-of-the-art image enhancement methods on MIT-5K test set.

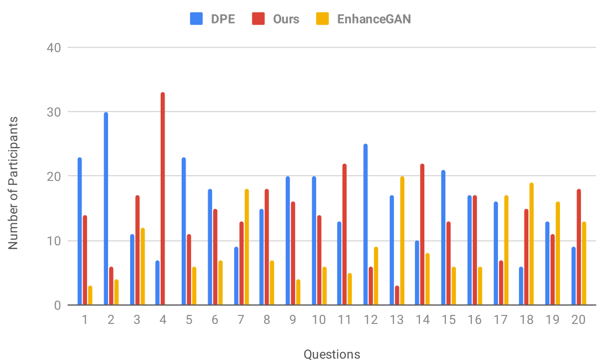

Figure 7. User study results, depicting the number of votes obtained per image.

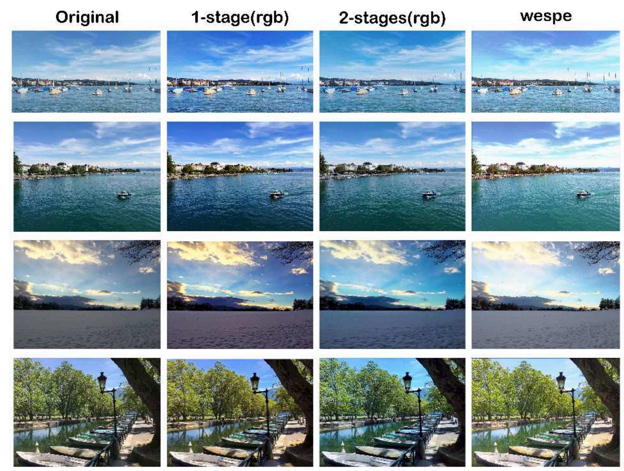

Figure 8. Comparison of our RGB models against WESPE [8]. From top to bottom: Original images from WESPE are taken *(from top to bottom)* with HTC One M9 Camera, Huawei P9 Leica Camera, iPhone 6 Camera and iPhone 6 Camera.

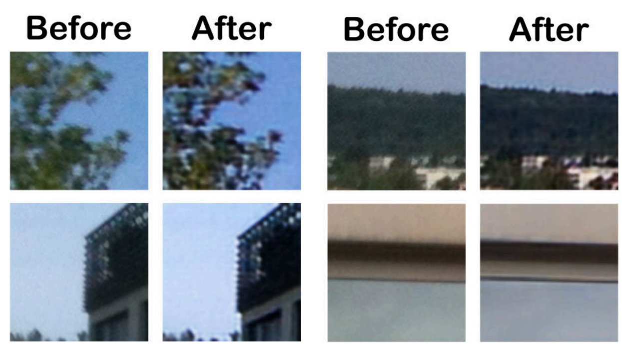

Figure 9. Visual results of our model on the image patches used in the NTIRE challenge

and CycleGAN.

Qualitative Evaluation. Due to the subjective nature of aesthetics, we conducted a user study to evaluate the qualitative performance of our models. We selected 20 random images from the MIT-5K test set and compared the images enhanced with our 1-stage RGB model to that using DPE and EnhanceGAN. Notably, as the MIT-5K dataset does not impose categories, the selected images are sampled from any categories including landscape, architecture, people, animal and flower. We recruited 40 participants for this study, which was hosted on the web browser. The results of the 20 images were placed side-by-side in a random order and participants were asked to choose one image that they think is the best enhanced image compared to its original state. Figure 7 shows the detailed results of the user study. Our 1-stage RGB model ranked second with 291 votes, losing marginally only to Deep Photo Enhancer (DPE) (323 votes) by 0.8 votes per participant while winning over EnhanceGAN (186 votes) by a much larger measure.

Figure 5 shows the visual comparison of enhanced images produced by our models versus that by DPE and EnhanceGAN. Notably, our model obtained higher scores for all five images (5/20), which are landscape and architecture images. On the other hand, it is not surprising that our model received lower votes for the two images in Figure 6, both of which contain a dominant subject of interest (person and flower). These types of images were not trained on our model. In summary, it is obvious that our model performed best for landscape and architecture images, but fail to obtain good scores for most of the image categories that our model was not trained on. In future, this could be addressed by employing a more diverse range of image categories for training.

Figure 8 compares our RGB models against the one enhanced in WESPE [8]. Visually, our results show more vibrant colors, with minimal alterations to the brightness levels. On the other hand, results generated by WESPE are generally brighter with only slight color enhancement. The increased brightness resulted in over-exposed effects in some of their results, particularly in the sky or cloud areas, e.g. WESPE's image at two middle rows.

Figure 10. Results of our model on the full images from the NTIRE dataset

	MOS	SSIM	PSNR	User on codelab
1	2.784	0.7907	22.3488	zxg1120101037
2	2.595	0.7912	22.4146	swz30
3	2.591	0.8029	22.4421	Sprite
4	2.554	0.7817	21.7388	BOE-IOT-AIBD
5	2.529	0.7876	22.1436	zhwzhong
6	2.527	0.7621	22.1654	iim lab
7	2.419	0.7822	21.7524	Jie_Liu
8	2.402	0.7189	21.3691	flavio
9	2.394	0.7840	21.9217	UltraVision
10	**2.293**	**0.7300**	**18.6852**	**nelson1996**
11	2.253	0.7569	18.3972	scape1989
12	n.a.	0.8	22.66	zheng222
13	very bad	0.7446	20.9666	KantiKumari

Table 3. Final results of NTIRE 2019 Image Enhancement Challenge.

5.2. NTIRE Challenge Dataset

For participation in the NTIRE Image Enhancement Challenge [9], we trained a 1-stage model in the HSV color space using the NTIRE dataset provided. The model was trained at an input resolution of 100×100 pixels. In the final evaluation phase, our model achieved an average PSNR score of 18.6852, an average SSIM score of 0.73 with an average perceptual quality (MOS) of 2.293 based on the release of the final results [9] depicted in Table 3.

By visually inspecting some of the 100×100 image patches tested (see Fig. 9, we observe that the enhanced outputs illustrate an improved sharpness and reduction in noise but they generally do not look too different from the input patches in terms of color vibrancy. This is likely due to our choice of submitting our HSV model instead of the stronger RGB model. Figure 10 shows the results of our approach on full images from the NTIRE dataset. Similarly, the enhanced images differ only slightly in terms of the color tones and overall brightness compared to the original images. We surmise that the shortcoming of our results on the NTIRE dataset could be due to the mismatch between the target domain of the NITRE dataset and our proposed training strategy— the challenge dataset was intended for standard image enhancement while our model is designed to learn aesthetically-driven enhancement based on from samples of aesthetically high and low images. The existing gap in our model to be addressed in our future work.

6. Conclusion

In conclusion, we propose a 2-way GAN method for aesthetic-based image enhancement based on CycleGAN, with several fine-grained modifications made to the original CycleGAN. We have conducted experiments on the impact of different color spaces; RGB, HSV and LAB, as well as training strategies with different number of stages. Our findings show that training performed in RGB color space reduces the amplification of noise significantly. The 1-stage training strategy demonstrates a good boost in saturation and contrast which could sometimes, result in oversaturation effects. Meanwhile, the 2-stages training strategy was able to solve the over-saturation problem, providing a less dramatic and more naturally enhanced image. Both qualitative and quantitative experiments were conducted to evaluate the performance of our proposed method against recent state-of-the-art approaches. Results show that our method is superior in terms of average PSNR score, validating our hypothesis on reducing amplification of noise. In terms of average SSIM score, our approach fared reasonably well though there is obvious room for improvement in future to address shortcomings in our method.

7. Acknowledgement

This work is supported in part by FRGS Research Grant No. FRGS/1/2018/ICT02/MMU/02/2, MMU Internal Grant No. MMUI/180153 and Multimedia University. We are grateful to the organizers of the NTIRE challenge who saw the merit of the contributions in this paper.

References

[1] Andrew P. Aitken, Christian Ledig, Lucas Theis, Jose Caballero, Zehan Wang, and Wenzhe Shi. Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize. *CoRR*, abs/1707.02937, 2017.

[2] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *Proceedings of IEEE CVPR*, 2011.

[3] Qifeng Chen, Jia Xu, and Vladlen Koltun. Fast Image Processing with Fully-Convolutional Networks. In *Proceedings IEEE ICCV*, 2017.

[4] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of IEEE CVPR*, pages 6306–6314, 2018.

[5] Y. Deng, C. C. Loy, and X. Tang. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, 34(4):80–106, July 2017.

[6] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In Yee Whye Teh and Mike Titterington, editors, *Proceedings of AISTATS*, volume 9, pages 249–256. PMLR, 2010.

[7] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks. *Proceedings of IEEE ICCV*, pages 3297–3305, 2017.

[8] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: Weakly supervised photo enhancer for digital cameras. 09 2017.

[9] Andrey Ignatov, Radu Timofte, et al. NTIRE 2019 Challenge on Image Enhancement: Methods and Results. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.

[10] Andrey Ignatov, Radu Timofte, Thang Van Vu, Tung Minh Luu, Trung X Pham, Cao Van Nguyen, Yongwoo Kim, Jae-Seok Choi, Munchurl Kim, Jie Huang, et al. Pirm challenge on perceptual image enhancement on smartphones: report. In *Proceedings of ECCV*, 2018.

[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 CVPR*, pages 1125–1134, 2017.

[12] Joon-Young Lee, Kalyan Sunkavalli, Zhe Lin, Xiaohui Shen, and In So Kweon. Automatic content-aware color and tone stylization. In *Proceedings of the IEEE CVPR*, pages 2470–2478, 2016.

[13] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. Saliency for image manipulation. *The Visual Computer*, 29(5):381–392, 2013.

[14] Nathan Moroney. Local color correction using non-linear masking. In *Color and Imaging Conference*, volume 2000, pages 108–111. Society for Imaging Science and Technology, 2000.

[15] Naila Murray, Luca Marchesotti, and Florent Perronnin. AVA: A large-scale database for aesthetic visual analysis. *Proceedings of IEEE CVPR*, pages 2408–2415, 2012.

[16] Tam V Nguyen, Bingbing Ni, Hairong Liu, Wei Xia, Jiebo Luo, Mohan Kankanhalli, and Shuicheng Yan. Image re-attentionizing. *IEEE Transactions on Multimedia*, 15(8):1910–1919, 2013.

[17] Andrea Polesel, Giovanni Ramponi, and V John Mathews. Image enhancement via adaptive unsharp masking. *IEEE Transactions on Image Processing*, 9(3):505–510, 2000.

[18] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13(1):100–111, 2004.

[19] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001.

[20] Alessandro Rizzi, Carlo Gatta, and Daniele Marini. A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters*, 24(11):1663–1677, 2003.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.

[22] Xiaoou Tang, Wei Luo, and Xiaogang Wang. Content-based photo quality assessment. *IEEE Transactions on Multimedia*, 15(8):1930–1943, Dec. 2013.

[23] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Proceedings of IEEE ICCV*, volume 98, page 2, 1998.

[24] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[25] Lai-Kuan Wong and Kok-Lim Low. Saliency retargeting: An approach to enhance image aesthetics. In *Proceedings of IEEE WACV*, pages 73–80. IEEE, 2011.

[26] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning*, pages 2048–2057, 2015.

[27] Zongben Xu, Xiangyu Chang, Fengmin Xu, and Hai Zhang. L-1/2 regularization: A thresholding representation theory and a fast solver. *IEEE Transactions on Neural Networks and Learning Systems*, 23:1013–27, 07 2012.

[28] Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. Automatic photo adjustment using deep neural networks. *ACM Transactions on Graphics (TOG)*, 35(2):11, 2016.

[29] J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of IEEE ICCV*, pages 2242–2251, 2017.