GENERATIVE ADVERSARIAL NETWORK USING WEIGHTED LOSS MAP AND REGIONAL FUSION TRAINING FOR LDR-TO-HDR IMAGE CONVERSION

Sung-Woon JUNG†, Hyuk-Ju KWON†, Dong-Min SON†, Nonmembers, and Sung-Hak LEE†∗, Member

SUMMARY High dynamic range (HDR) imaging refers to digital image processing that modifies the range of color and contrast to enhance image visibility. To create an HDR image, two or more images that include various information are needed. In order to convert low dynamic range (LDR) images to HDR images, we consider the possibility of using a generative adversarial network (GAN) as an appropriate deep neural network. Deep learning requires a great deal of data in order to build a module, but once the module is created, it is convenient to use. In this paper, we propose a weight map for local luminance based on learning to reconstruct locally tone-mapped images.

key words: CycleGAN, HDR, image conversion, weighted loss function, regional fusion training

1. Introduction

Recently in the field of deep learning, the use of generative adversarial networks (GANs) for image generation has attracted considerable attention [1]. In particular, the CycleGAN architecture shows efficient image conversion with a small amount of data compared with other GAN models. The original GAN comprised a generator and a discriminator [1]–[3], [6]. In this design, the generator creates an image using latent space in the feature map, while the discriminator employs a convolutional neural network (CNN) structure [7] to determine whether a generated image is real or image [1], [8]. To decrease the GAN’s overall loss value, the loss value of the discriminator should be increased while the generator’s loss value should be decreased. Because of this competitive structure, the GAN can produce a generated image that is closer to reality by reducing the generator loss.

In this adversarial network model, pix2pix which the image-to-image translation method is trained using target and reference images to form a single module [2], which attempts to convert the input image into the style of the reference image. However, converting low dynamic range (LDR) images to high dynamic range (HDR) images using pix2pix requires large amounts of image data and training time. Rendered HDR images using pix2pix illustrate undesirable results such as unrelated colors and boundary errors (see Fig. 1 for an example of this). The CycleGAN architecture can mitigate the disadvantages of pix2pix, showing efficient image conversion using a small amount of data compared with other GAN models [3]. CycleGAN incorporates a reconstructive approach based on the concept of cycle consistency loss, which calculates whether a fake image is properly restored to the original image. This structure allows training with a small amount of unpaired data. In this paper, we introduce an enhanced CycleGAN-based method that consists of increasing the exposure of dark images, maintaining the color map of the original image, and toning up the image locally while maintaining the bright part. We confirm that the proposed method is faster and more efficient than existing methods.

2. Related Work

2.1 Generative Adversarial Networks

CNNs offer a learning method that deduces correct output through decomposition and analysis of data by the neural network. In GANs, by contrast, the discriminator is an auto-encoder with a CNN structure that differentiates real and fake images provided by the generator, while the generator creates new results to attempt to fool the discriminator [2]. The auto-encoder converts the image while creating a feature map through the CNN and restoring it again. In order to reduce the generator’s loss value, the performance of the discriminator should be fixed; otherwise, the discriminator overpowers the generator’s performance. This structure creates unpredictable results because it incorporates a form of unsupervised training, making it difficult to produce a result within detailed constraints [5].

2.2 Image-to-Image Translation

A practical approach using a GAN structure is offered by pix2pix. The intention is to work with two classes of images, A and B, such that image A is chosen to be similar to image B [2]. However, this method requires images to be associated with each other through training in pairs. For instance, in order to convert a natural photographic image
into a Monet painting style, a large amount of data drawn directly from the Monet style data set is required for style description [3].

In the CycleGAN approach, unlike pix2pix, machine learning of unrelated pairs of images is possible. In fact, “reconstruction” refers to restoration by the generator back to the original. CycleGAN uses the same generator function as pix2pix, but in the opposite direction. Compared to pix2pix, smaller data sets are effective for coloring objects within the images, while boundaries are also relatively well preserved. However, a major issue with this approach is faulty colorization.

2.3 Residual Dense Networks

Residual dense networks (RDN) use all the feature maps derived from the learning process by computing past learned information with the current information and accumulating the calculation results [4]. This method has shown great efficiency in restoring detail. In this paper, we apply this RDN approach to the CycleGAN architecture to improve learning speed and adjust learning rate.

3. Proposed Method

3.1 Luminance Channel Training

In previous GAN approaches, RGB channels for all images are used for learning. Therefore, the generator considers all areas of all reference images. In LDR-to-HDR image translation learning, however, unnecessary color information can be neglected and learning time can be reduced by requiring a lower dimension than the existing method. Within presented algorithm, only the luminance channel from Lab color space is used. When training, only the luminance channel L is trained, while the a and b color channels are preserved.

3.2 Weighted Loss Maps (WLM)

For dark-area learning in the L channel of LDR images, we derive a weighted loss map \( W_B \) from the inverse Gaussian map in the loss function \( L \) for the target and reference images. The loss function then concentrates on the dark regions of the image using \( W_B \). The modified loss function is as follows:

\[
W_B = 1 - N(F(B)) \tag{1}
\]

\[
L = \|W_B(A - B)\|_1 \tag{2}
\]

where a Gaussian filter of size \( 11 \times 11 \) for a blurred or LDR reference image \( B \) is expressed as \( F(B) \); the image is then normalized by \( N() \) and reversed.

The learning growth rate can be faster than the existing method in Eq. (2). In fact, some improved results were obtained over 200 data sets at as few as 30 epochs (check in Fig. 4). We observed that the existing method cannot maintain the original colors in the same epochs.

3.3 Regional Fusion Training (RFT)

The weight map proposed in Eqs. (3) and (4) improves the weighting for the LDR image dark area training; the rendered HDR image results show that dark areas in the LDR image are effectively enhanced in intensity while oversaturation occurs in the bright areas of the LDR image. To prevent oversaturation in the image, it is necessary to change the learning rate of the generator locally and reconstruct the image according to the brightness of the reference image. We designed that the generator at state \( N \) is evolved from the previous state at \( N-1 \) and current state \( N \) using weight map \( W_B \). It can make the bright area is more train than dark area using previous data and accelerate learning rate. The equation is shown in Eq. (3). The reconstruction is modeled by an inverse function corresponding to the generator. Therefore, in Eq. (4) the scheme of reconstruction is opposite to that of the generator. The proposed generator and reconstruction are as follows:

\[
G_n = W_B \times G_{n-1} + (1 - W_B) \times G_n \tag{3}
\]

\[
R_n = W_A \times R_n + (1 - W_A) \times R_{n-1} \tag{4}
\]

where \( G \) represents a fake image created by the generator, \( R \) is the reconstructed image, and \( W_A \) and \( W_B \) are weight maps for the reference and target images respectively. The current training result is denoted \( n \), while \( n-1 \) indicates the previous training result.

Figure 2 shows an overview of the algorithm describing the proposed method. The red and blue boxes outline the

![Flow chart of the proposed method](image-url)
training and testing processes respectively. When the module is created, only the luminance channel is extracted from the input image and passed into the module. Chrominance channels $a$ and $b$ of the input image are preserved. The output image is composed using the chrominance channels of the input image and the trained luminance channel.

4. Simulation Results

The proposed method was simulated in two ways depending on the data set configuration. In the first configuration, 502 training data sets of $256 \times 256$ HDR and LDR images were used. The images were sourced from Mark Fairchild’s “HDR Photographic Survey” [9] and the total data sets number 1004. One hundred kinds of bright natural images not used for training were used as the validation data set. All data sets were paired with images associated with each other. In the second data set configuration, experiments were performed using training data set images with EVs (exposure values) set to $-3$, $-4$, and $-5$. The first method is an intuitive training approach for LDR-to-HDR image conversion. The second training method transforms dark images to bright images.

Figures 3 (a) and (b) show the data from the first and second experiment, respectively, with the LDR and HDR image pair. The second experiment data uses dark and light images with EVs of $-3$, $-4$, and $-5$. The red and blue boxes outline the target and reference images. The input image of each method is the LDR image in Fig. 3 (a). Figures 3 (c) and (d) show the results of CycleGAN, luminance CycleGAN (Lum GAN), and the proposed method for each of Figs. 3 (a) and (b). CycleGAN uses RGB images for training, while Lum GAN and the proposed method use only the luminance component. When comparing the results of the first and second experiments, the EV data set’s results shown in Fig. 3 (d) are brighter than the HDR data set’s results in Fig. 3 (c).

The Lum GAN image is brighter than the previous CycleGAN; however, there is noise in the picture. In Fig. 3 (c), the color distortion and noise in the CycleGAN image can be clearly seen. In the Lum GAN image, the color distortion is reduced but noise still exists. Finally, the results of the proposed method have better clarity and are closer to the HDR image in Fig. 3 (b) than the results of the other methods.

Figure 4 compares the proposed method with Lum GAN using the LDR and HDR image pair data set.

Figure 4 (a) shows the Lum GAN results while Fig. 4 (b) shows the results of the proposed method. Each set of results shows the progress of training according to epochs. In Fig. 4 (a), with 30 epochs, the training does not accomplish the desired results. In Fig. 4 (b), the proposed method shows good results with fewer epochs than Lum GAN (Fig. 4 (a)). Even though the proposed method employs fewer epochs, it exhibits a high learning rate. However, the proposed method’s batch speed is relatively slow, since the RFT stores previous epochs’ information.

In Fig. 5, LDR and HDR images are compared with the rendered output images using CycleGAN, Lum GAN, and proposed method respectively. HDR images are well represented in the dark region while maintaining the color components of the LDR image. Each result image can be compared in terms of local detail and color representation performance.

In Fig. 5 (e), the rendered global tone, local detail, and color becomes clear and close to the HDR targets. But
in Fig. 5(c) and (e), the brightness of the image is partially saturated and some boundaries are difficult to distinguish. To compare the imaging results quantitatively, we adopted three image-quality metrics for evaluations. The HDR-VDP-2 as an HDR visual detection predictor and the MS-SSIM based on multi-scale structural similarity are used for HDR image visibility and quality assessments [10], [11]. Additionally the PSNR (peak signal-to-noise ratio) is used as a representation fidelity measurement between the result images and the target images. Comparative charts are presented in Fig. 6. Excepting some of the scores, on average, the performance of the proposed method is better than those...
of the others. In particular, it is confirmed that the performance improvement compared with the evaluation of visibility and structural similarity is large.

5. Conclusions

In this paper, we propose a weight map for local luminance-based learning to reconstruct locally tone-mapped images. When converting LDR to HDR, the proposed method is faster and more efficient than the existing CycleGAN method, but does have disadvantages when comparing image results the actual HDR image. For example, in oversaturated areas, the proposed method tends to lose some boundary information and also produces slightly darker images than the actual HDR images. Further research is needed to overcome information loss in oversaturated areas and to raise the quality of the results to be closer to the quality of the HDR images.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) and the BK21 Plus project funded by the Ministry of Education, Korea (NRF-2019R1D1A3A03020225, 21A20131600011).

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Advances in Neural Information Processing Systems (NIPS), pp.2672–2680, 2014.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A.A. Efros, “Image-to-image translation with Conditional Adversarial Networks,” CVPR, pp.1125–1134, 2017.

[3] J.-Y. Zhu, T. Park, P. Isola, and A.A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” International Conference on Computer Vision (ICCV), to appear, pp.2223–2232, 2017.

[4] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual Dense Network for Image Super-Resolution,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2472–2481, 2018.

[5] G. Eilertsen, J. Kronander, G. Denes, R.K. Mantiuk, and J. Unger, “HDR image reconstruction from a single exposure using deep CNNs,” ACM, vol.36, no.6, pp.1–15, 2017.

[6] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” arXive:1609.04802v5, 2017.

[7] P.Y. Simard, D. Steinkraus, and J.C. Platt, “Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis,” ICDAR, pp.958–962, 2003.

[8] B. Shi, X. Bai, and C. Yao, “An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.39, no.11, pp.2298–2304, 2017.

[9] http://rit-mcsl.org/fairchild/HDR.html

[10] R. Mantiuk, K.J. Kim, A.G. Rempel, and W. Heidrich, “HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions,” ACM Transactions on graphics (TOG), vol.30, no.4, pp.1–14, 2011.

[11] Z. Wang, E.P. Simoncelli, and A.C. Bovik, “Multiscale structural similarity for image quality assessment,” Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers, pp.1398–1402, Nov. 2003.