Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Low Illumination Enhancement For Object Detection In Self-Driving*

Yangyang Qu[1,2,3,4], Yongsheng Ou[1,3,4], and Rong Xiong[1,3,4,†]

***Abstract*— Object detection plays an important role in the field of self-driving. Illumination has a great impact on object detection, but most of the current methods do not solve the problem of object detection in poor light environment well. We propose a network that can optimize the image conversion which is based on the Cycle Generative Adversarial Networks (CycleGAN). We redesign the discriminator network of CycleGAN, add additional discriminators, optimize multiple parts of the network such as loss functions, and add object detection networks after converting the network. The Robot Car dataset of Oxford University is utilized to verify the effectiveness of the proposed method, and the results proved that our method can significantly improve the detection accuracy and increase the detected object number in low illumination environment.**

***Index Terms*— GAN, Image Enhancement,self-driving,low illumination**

I. INTRODUCTION

Perception in self-driving technology plays an important role in the process of self-driving. There are three basic approaches currently being adopted in research into Perception research. It contains the lidar-based method, lidar-based plus camera method and the camera-based method [1]. At present, visual perception is indispensable in the object recognition technology of automatic driving. The camera can bring a larger information volume, and users can better understand the driving environment. In the more complex traffic environment such as the urban environment, these details become very important. However, the performance of the camera is greatly affected by the light and some other elements. During the day, the camera can easily sense changes in the surrounding environment, but at night because the light is not enough, Images taken under low light conditions will suffer from low contrast, low visibility and high ISO noise, so the research on how to improve the self-driving cars for night vision image perception, of self-driving cars at night driving safety is of great significance. Traditional methods have made some progress in processing dark night images, but they still have poor effects on background, light interference, rain, snow and other bad weather. Moreover, the time consumption of traditional methods is too long to meet the demand of real-time. The main problem of this paper is the object detection in the case of poor light environment. To cope with the problem of poor object detection performance under low light conditions, this paper proposes a new solution, which converts the night scene images into daytime images for object detection by generating antagonistic networks without losing the objects as many as possible. The main contributions of this paper are:

(1) Our network structure successfully adopts the method of unpaired training to train images. This training strategy eliminates the dependence on paired training data and makes our method more generalized in testing.

(2) By adding colour and edge detectors, our network structure makes the results have a good effect on dark light enhancement of both black and white images and colour images. Compared with other methods, our network achieves better results through multiple sets of comparative experiments.

(3) We add the object detection network to the Cycle Generative Adversarial Networks network, and get a good effect. The experimental results prove that our method can improve the object detection effect in the environment with poor light.

II. RELATED WORK

Traditional image processing is an early image conversion method, but the effect performance is poor limited. After the emergence of deep learning method was developed, it had been widely applied in image conversion such as using the Convolutional Neural Networks (CNN)framework [2], [3].

A. Image conversion

Generative Adversarial Networks (GAN), a network structure with inherently strong ability in image generation, using the GAN method to do image conversion has become a relatively popular method [4]. AugGAN is a very good network structure that can detect objects under poor light [5], but it dependents on segmentation tag. If the data is

[1] The authors are with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. yy.qu, rong.xiong, ys.ou@siat.ac.cn

[2]University of Chinese Academy of Sciences, Beijing 100049, China.

[3]Guangdong ProvincialKey Lab of Robotics and Intelligent System, ShenzhenInstitutes of Advanced Technology, Chinese Academy ofSciences.

[4]CAS Key Laboratory of Human-MachineIntelligence-Synergy Systems, ShenzhenInstitutes of Advanced Technology. Rong Xiong is the corresponding author.

†Rong Xiong is the corresponding author rong.xiong@siat.ac.cn

*This work was jointly supported by National Natural Science Foundation of China (Grants No. U1613210), Guangdong Special Support Program (2017TX04X265), Primary Research & Development Plan of Guangdong Province (2019B090915002), and Shenzhen Fundamental Research Program (JCYJ20170413165528221).

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

sufficient, the availability of AugGAN's converted data is very high. Pix2pix which requires pairs of images [6] is also a good network which can convert the night image to day well. GAN consists of two subnetworks, G (Generator) and D (Discriminator). G is used to generate the image. After receiving a random noise z, it generates the image G(z). D checks whether the generated image G(z) is true, and the output D (x) represents the probability that x is true. In training process, G's goal is to generate as many real pictures as possible to deceive the discriminant network D. The goal of D is to try to separate the images generated by G from the real ones. Thus, G and D constitute a dynamic "game process". Training GAN requires two losses: the reconstruction loss of the generator hopes to generate G(G(x)) as similar as possible to the original picture x. The discriminator discriminates Loss, and the generated false picture and the original true picture are input into the discriminator to classify the Loss:

$$\begin{aligned}\min_{G}\max_{D} V(D,G) = E_{x\sim P_{data(x)}}\left[\log\left(D\left(x\right)\right)\right] \\ +E_{z\sim P(z)}\left[\log\left(1-D\left(z\right)\right)\right]\end{aligned} \tag{1}$$

CycleGAN network introduces unsupervised learning into image conversion. CycleGAN is essentially two symmetrical GAN, forming a ring network. The two GAN shares two generators, each with a discriminator. After this, Horia Poravd added additional periodic losses to the original CycleGAN setting[7],they proposed a "circular consistency loss":

$$\begin{aligned}L_{LSGAN}\left(D_Y, G, X, Y\right) = E_{y\sim P_{data(y)}}\left[\left(D_y\left(y\right)-1\right)^2\right] \\ +E_{x\sim P_{data}(x)}\left[\left(1-D_Y\left(G\left(x\right)\right)\right)^2\right]\end{aligned} \tag{2}$$

It could also achieve a better image from night to day. And their full objective is:

$$\begin{aligned}L\left(G,F,D_X,D_Y\right) = L_{LSGAN}\left(D_Y,G,X,Y\right) \\ +L_{LSGAN}\left(D,F,X,Y\right)+\lambda L_{cyc}\left(F,G\right)\end{aligned} \tag{3}$$

TodayGAN network[8] which modifies ComboGAN network[9] achieves a good image conversion. But their main job is the The target detection algorithm based on deep learning needs no manual features to design map matching, which is different from ours. Our work is also based on unsupervised learning for image conversion. The loss of the discriminator network is modified on the basis of CycleGAN network, and a variety of constraints are added to realize image conversion.

B. Object detection

Object detection is the basis of many computer vision tasks. The object detection algorithm based on deep learning needs no manual feature design, and it has good feature expression ability and excellent detection accuracy. It has surpassed the traditional detection methods and becomes the mainstream of object detection.

At present, the classification of object detection mainly includes two types. The first type is object detection based on regional nomination. Regions with CNN feature (RCNN) breaks through the traditional idea of the object detection algorithm [10], which is the first successful breakthrough of deep learning in the field of object detection. Based on RCNN, Fast-RCNN adopts adaptive scale-pooling to optimize the entire network, which avoids redundant feature extraction operations in RCNN and improves the accuracy of network identification [11]. Faster RCNN further improves the speed by constructing regional recommendation network to extract candidate boxes[12]. Mask-RCNN adds a Mask Prediction Branch on the basis of Faster-RCNN and proposed ROI Align, further improving the effect [13]. The second method is object detection based on end-to-end learning. This method does not need to extract candidate regions in advance. You only look onec (YOLO) simplifies the whole process of object detection, integrates object determination and recognition, and greatly improves the running speed [14]. Single Shot MultiBox Detector (SSD) uses RPN in multiple feature layers of CNN before classification and border regression [15] , and the detection of small objects in the original image can also have more accurate detection results. Compared with the previous two generations, YOLO V3 uses the multi-scale prediction method and the better basic classification network darknet-53 and classifier. The accuracy of YOLO V3 is similar to that of SSD, but three times faster than SSD.

C. Image Enhancement

Low Illumination enhancement is always considered as an important part of image problem. Classical methods include adaptive histogram equalization (AHE) [16], multi-scale Retinex model [17]. Some researchers noticed the similarity between defogging and dark light enhancement in the image, so they defogged the image. Recently,[18] proposed a luminance pass filter to decompose the image into reflectivity and illumination to determine the detail and naturalness of the image respectively. And a double logarithmic transformation is proposed, which is used to map light to strike a balance between detail and nature. [19]focuses on another work to estimate a coefficient matrix of the enhanced illumination map by addition. This approach produced impressive results.[20] a joint low-light image enhancement and noise reduction model based on continuous image sequence decomposition was proposed. [21] designed a multi-exposure fusion framework for low-light image enhancement. Based on the framework, they propose a dual-exposure fusion algorithm to provide an accurate contrast and lightness enhancement.

III. PROPOSED METHORD

The main idea of realizing night recognition is to convert images from night to day through CNN network. The

TABLE I

NETWORK ARCHITECTURE FOR THE IMAGE-TO-IMAGE TRANSLATION EXPERIMENTS.

Layer	**ResGenEncoders**	**Layer Info**		
		Convolution Output	**Kernel Size**	**Stride**
1	CONV,ReLU	64	7	1
2	CONV,ReLU	128	3	2
3	ResnetBlock,ReLU	256	3	1
4	ResnetBlock,ReLU	256	3	1
5	ResnetBlock,ReLU	256	3	1
6	ResnetBlock,ReLU	256	3	1
Layer	**ResGenDecoders**	**Layer Info**		
		Convolution Output	**Kernel Size**	**Stride**
1	CONV,ReLU	256	3	1
2	CONV	256	3	2
3	ResnetBlock,ReLU	256	3	1
4	ResnetBlock,ReLU	256	3	1
5	ResnetBlock,ReLU	256	3	1
6	ConvTransposed,ReLU	128	4	2
7	ConvTransposed,ReLU	64	4	2
8	ConvTransposed,ReLU	3	7	1
Layer	**Discriminator1,2**	**Layer Info**		
		Convolution Output	**Kernel Size**	**Stride**
1	CONV,ReLU	64	4	2
2	CONV,ReLU	128	4	2
3	CONV,ReLU	256	4	2
4	CONV,ReLU	512	4	2
5	CONV,ReLU	1	4	2

previous methods mainly rely on the traditional image enhancement methods. After the emergence of deep learning, the CNN network and the improved correlation algorithm can also realize the related methods.The first part is the transformation network. The idea behind transformation networks is the same as CycleGAN, where two networks are pitted against each other to produce a picture of the object.

The first part is the transformation network. The first network is a generator network for converting images into object images. The structure of the generator network is similar to that of the style transformation network[2]. It contains an encoder, converter, and decoder three parts. The first step we use CNN to extract features from the input image. After passing through the encoder network, the image is compressed into 256*64*64 feature vectors. The converter network is composed of four Reset blocks [22],each Reset module is composed of two convolution layers, two normalization layers, and an activation function. Part of the data in the converter network is converted by the conversion network, and the other part of the data will be directly converted to the output without conversion. This ensures that the previous input data information directly ACTS on the later network layer, narrowing the deviation, otherwise the output may deviate from the object contour. The decoding network is completely opposite to the encoding network, and the deconvolution network is used to restore features from feature vectors and generate object images. The structure is so that the object image can be generated. After sampling down and up, the generator network can reduce the computation and be beneficial to acceptable receptive fields. The generator is used to reconstruct the image, with the purpose of making the image as similar as possible to the original image. The loss of the generator is expressed as follows:

$$L_G = E(G_2(G_1) - real) \tag{4}$$

The Discriminator network is designed to be an innovation of our approach, which is used to determine whether the image just generated is true or false. Discriminator network is a modification of PatchGAN which is widely used in generation antagonism network. There are two generators and two discriminators of the same structure. Each epoch is updated with data from the network of generators and discriminators. Then there are volumes and networks for extracting features from images. The network structure consists of four Sequential layers, each contains a convolutional layer and a Sequential activation layer. Similar to [6], we apply two

discriminators on the output of each generator to minimize the following loss: D1 and D2 on the output of generator G1, and D3 and D4 on the output of generator G2. The discriminator network contains two parts, D1 is to detect the edge of the image and the D2 is to detect the change of the grey level of the image. This loss is formulated as:

$$L_{D11} = (D_1 - 1)^2 + (D_1(G_1))^2 \tag{5}$$

$$L_{D12} = (D_2 - 1)^2 + (D_2(G_1))^2 \tag{6}$$

$$L_D = \lambda_1 L_{D11} + \lambda_2 L_{D12} + \lambda_1 L_{D21} + \lambda_2 L_{D22} \tag{7}$$

In the final calculation of the loss, the average value of each loss function is calculated to obtain better optimization results. Thus, the generation network and the authentication network have been constructed. Detailed architecture of our network is given in Table 1. The Loss of loop consistency is Smooth L1 Loss L_cyc, which is more stable after conventional L1 Loss in training.

GAN network loss function using least squares loss, commonly used cross entropy loss is different, the cross-entropy loss will not go to think is really network optimization that has been identified may lead to optimizing image quality is not high (take a picture), the least squares also includes to generate network will be far away from the decision-making border generated images generated in the direction of the decision boundary [23]. Loss includes loss of generation network adversarial loss, which aims to make the generated data close to the real data distribution. In addition, cycle consistency loss is included to ensure that the samples generated by the two generating networks are consistent.The losses generated by the network include those above. The loss of the identification network is the sum of two detection network losses. And the objective function is given as follows:

$$L_{all} = L_D + L_{cyc} + L_G \tag{8}$$

At the end of the above steps, we connected the YOLO V3 object detection network behind the image generator network to detect objects in the daytime image generated in the previous step.

IV. TRAINING

A. The dataset

The dataset we used is the Oxford Robot Car dataset [24]. It contains more than 20 million images of 1024 x 1024 resolutions. We select More than 30,000 pictures of them as the training set. It contains half of the night, half of the day. Since our goal is to convert the night image into the day image, our test set only shows the night image, but we also put part of the night and rainy photos in the test set, so we can detect the robustness and portability of the system. Because our final purpose is to carry out object detection, we choose scenes with more complex scenes.

B. Training

80% of our dataset is used as a test set and 20% as a training set. The training set images are divided into two parts. The first folder contains photos of the night, and the second folder contains photos of the day. And then we need to initialize all losses. We adopted Adam solver and set the initial learning rate of 0.0002 and the ramda as [9]. To add to the controlled trials, we also trained the CycleGAN network. In CycleGAN network training, set the same learning rate as our own designed network and learn the same epoch. The results are more comparable. At the same time, when we train black and white images and colour images, we can verify whether our network has a good conversion effect for both black and white images and colour images. According to the training results, we tested the night image conversion effect and image detection results of our designed network structure respectively. Since the data set is a dark data set, and the data set is not labeled, what we calculate is whether the confidence level has been improved and whether the number of detected objects has been improved. We also tested night images on the CycleGAN network.

V. EXPERIMENTS

To test our network image enhancement, we chose several networks that are currently used in comparative experiments. Including the new method published in 2017 and 2018 (we adopted AMSR, DONG, LINE, Ying, multiscale-Retinex, BIMEF, NPE), the results show that this particular scene image enhancement effect is not obvious in self-driving. As shown in Fig.1, most of the network structures have poor dark light enhancement effect on the automatic driving image.

By contrast, our network architecture has produced remarkable results.Fig.2 shows how well our network structure works in black and white. The contrast of other network structures to the black and white image enhancement effect is very not obvious. The results show that our method has better robustness and adaptability.

Fig.3 shows the object detection effect of our network transformed images through the YOLO V3 network. We found that the results of object detection were significantly improved. Table 2 shows the test results of our test result. The table shows that the number of objects detected by the test set and the improvement of confidence.Both of them have been greatly improved when the same objects are detected. We randomly select the contents of 100 images for calculation. In the randomly selected test set, the performance of object detection is significantly improved. The number of objects detected tripled, and the confidence is increased by 21%. In general, compared with the object detection effect of the original images in the dark scene, our solution has greatly improved the object detection effect.

Fig. 1. Comparison with other methods by using the RGB images, First column: the original image; Second column: the results of AMSR; 3rd column: the results of Dong; 4th column: the results of LIME; 5th column: the results of NPE; 6th column: the results of BIMEF; 7th column: the results of multiscale Retinex; 8th column: the results of $Ying\_2017\_ICCV$; 9th column: the results of Our methods.

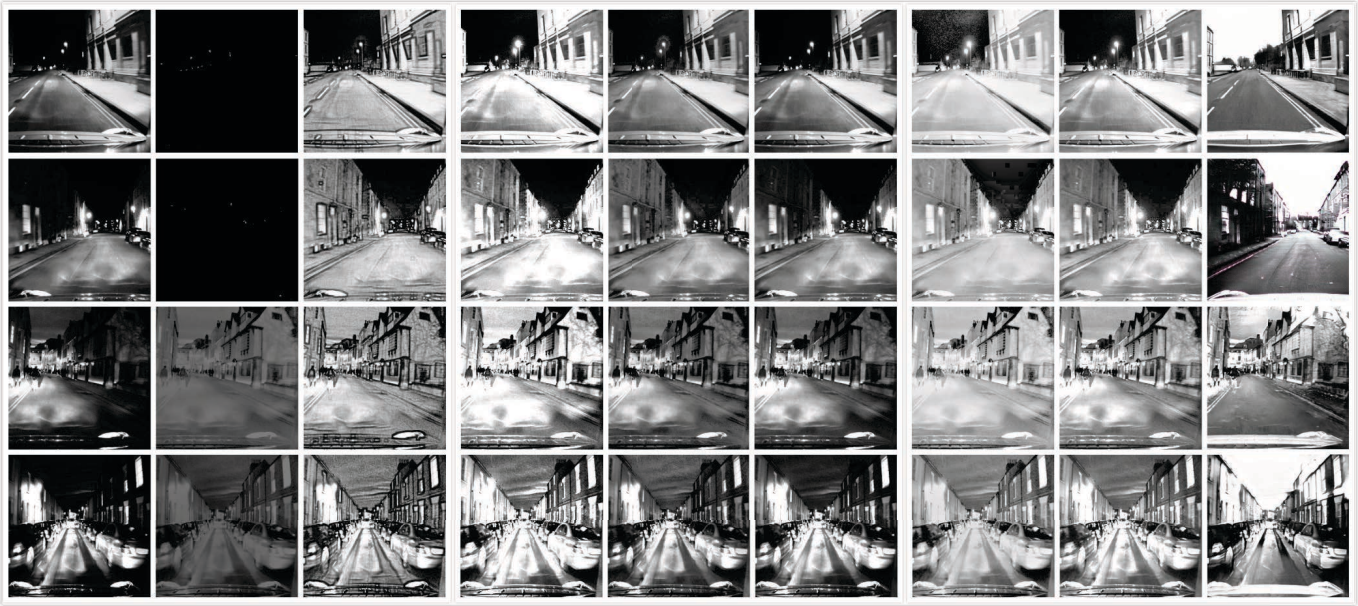

Fig. 2. Comparison with other methods by using the grayscale image images, First column: the original image; Second column: the results of AMSR; 3rd column: the results of Dong; 4th column: the results of LIME; 5th column: the results of NPE; 6th column: the results of BIMEF; 7th column: the results of multiscale-Retinex; 8th column: the results of $Ying\_2017\_ICCV$; 9th column: the results of Our methods.

VI. CONCLUSION

In this paper, we presented and experimentally validated a method which can improve the object detection effect by image transformation in the scene with poor light. The main contributions of this paper are to design a network which can improve the image transformation performance for object detection. We redesigned a network that included CycleGAN network and another discriminator network. Besides, we added a object detection network on the basis of our network and get good results. The experimental results show that the number of detected objects and the detection accuracy has been greatly improved. It can be proved that our method

Fig. 3. The image shows the object detection results of the original image as well as the image transformed by our trained transformation network after passing through the YOLO V3 network, First column: the original image; Second column: the results of AMSR; 3rd column: the results of Dong; 4th column: the results of LIME; 5th column: the results of NPE; 6th column: the results of BIMEF; 7th column: the results of multiscale Retinex; 8th column: the results of $Ying\_2017\_ICCV$; 9th column: the results of Our methods.

TABLE II

THE NUMBER OF DETECTED OBJECTS AND DEGREE OF CONFIDENCE

	Number of objects detected	**Degree of confidence**
Original	96	0.6652
Ours	162	0.8785

can achieve good object detection performance in the dark environment. Although our network has improved the object detection in the dark scene, if the light in the image is too strong, or the object is too close to the background, the objects still can be hard to be detected. In the next stage, we will continue to optimize the existing framework, integrate the transformation of other complex scenarios, and solve the object detection task in multiple scenarios. In addition, the future work will focus on optimizing the data set of YOLO network to improve the detection effect.

REFERENCES

[1] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. F. R. Jesus, R. F. Berriel, T. M. Paixão, F. W. Mutz, T. Oliveira-Santos, and A. F. de Souza, "Self-driving cars: A survey," *CoRR*, vol. abs/1901.04407, 2019. [Online]. Available: http://arxiv.org/abs/1901.04407

[2] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Computer Vision & Pattern Recognition*, 2016.

[3] Y. Shih, S. Paris, F. Durand, and W. Freeman, "Data-driven hallucination of different times of day from a single outdoor photo," *ACM Transactions on Graphics (TOG)*, vol. 32, 11 2013.

[4] J. Zhang, Z. Fan, G. Cao, and X. Qin, "St-gan: Unsupervised facial image semantic transformation using generative adversarial networks," , p. 598, 2017.

[5] S. W. Huang, C. T. Lin, S. P. Chen, Y. Y. Wu, P. H. Hsu, and S. H. Lai, "Auggan: Cross domain adaptation with gan-based data augmentation," 2018.

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

[7] H. Porav, W. Maddern, and P. Newman, "Adversarial training for adverse conditions: Robust metric localisation using appearance transfer," 2018.

[8] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. V. Gool, "Night-to-day image translation for retrieval-based localization," *CoRR*, vol. abs/1809.09767, 2018. [Online]. Available: http://arxiv.org/abs/1809.09767

[9] A. Anoosheh, E. Agustsson, R. Timofte, and L. V. Gool, "Combogan: Unrestrained scalability for image domain translation," *CoRR*, vol. abs/1712.06909, 2017. [Online]. Available: http://arxiv.org/abs/1712.06909

[10] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *CoRR*, vol. abs/1311.2524, 2013. [Online]. Available: http://arxiv.org/abs/1311.2524

[11] R. B. Girshick, "Fast R-CNN," *CoRR*, vol. abs/1504.08083, 2015. [Online]. Available: http://arxiv.org/abs/1504.08083

[12] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: http://arxiv.org/abs/1506.01497

[13] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," *CoRR*, vol. abs/1703.06870, 2017. [Online]. Available: http://arxiv.org/abs/1703.06870

[14] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015. [Online]. Available: http://arxiv.org/abs/1506.02640

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," *CoRR*, vol. abs/1512.02325, 2015. [Online]. Available: http://arxiv.org/abs/1512.02325

[16] PIZER, M. S., AMBURN, P. E., AUSTIN, D. J., CROMARTIE, and GESELOWITZ, "Adaptive histogram equalization and its variations," *Computer Vision Graphics & Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[17] D. J. Jobson, . Rahman, Z., and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 2002.

[18] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, 05 2013.

[19] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Trans Image Process*, vol. 26, no. 2, pp. 982–993, 2017.

[20] X. Ren, M. Li, W. H. Cheng, and J. Liu, "Joint enhancement and denoising method via sequential decomposition," 2018.

[21] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," *CoRR*, vol. abs/1711.00591, 2017. [Online]. Available: http://arxiv.org/abs/1711.00591

[22] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018. [Online]. Available: http://arxiv.org/abs/1804.02767

[23] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang, "Multi-class generative adversarial networks with the L2 loss function," *CoRR*, vol. abs/1611.04076, 2016. [Online]. Available: http://arxiv.org/abs/1611.04076

[24] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Research*, vol. 36, no. 1, p. 0278364916679498, 2016.