e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:03/March-2022 Impact Factor- 6.752 www.irjmets.com

NIGHT TO DAY AND DAY TO NIGHT IMAGE TRANSFER USING GENERATIVE ADVERSARIAL NETWORK

Binod Adhikari*1, Hari K.C.*2, Sharan Thapa*3

*1,2,3IOE Pashchimanchal Campus , Tribhuvan University, Nepal.

ABSTRACT

Image translation, transformation and augmentation is one of the most state of the art research topics in recent time and will continue for long future for more advancement. With the far and wide reaching of machine learning and artificial intelligence in daily life the image translation is vital for automatization of everyday life specially for night time driving. Many robotics applications and artificial intelligence application are seeking efficient algorithm and diverse dataset to perform end to end learning. More realistic translation of images for adverse lighting condition will lead towards more precise accurate and useful day to day application like night time driving, object detection, image enhancement, improved visual localization and many more. This research work aims to purpose a method for providing high quality transformation of night time image to day time image. Cycle Generative Adversarial Network employed in image processing is expected to provide better result outperforming other algorithms. Cycle Generative Adversarial Network with U-NET as generator and PatchGAN as discriminator trained to minimize the identity loss during image generation. Translated image of night time for corresponding day time image of same location resulted structural similarity of 78 percentage.

Keywords: Image Transformation, Cycle GAN, Night To Day, Low Light.

I. INTRODUCTION

Machine learning can be simply understood as field of computer science which gives computers the ability to learn without explicit programming. Sub field of machine learning: deep learning algorithms attempts to learn multi layer representation in hierarchical manner. Algorithms that implements multi-layer hierarchical relations are state of the art algorithms for image translation from images. Detecting objects from image and image translation has history of more than two decades and still searching for easy to implement and accurate detection and translation methods in both general translation and translational and detection under specific application scenarios. The use of image translation is in much real world application such as: optical character recognition, self driving cars, tracking objects, face detection and face recognition, identity verification, object extraction from an image or video, smile detection, activity recognition, pedestrian detection, digital watermarking, medical imaging, ball tracking in sports, object recognition as image search, manufacturing industries, robotics, automated cctv, automated image annotation, automatic target recognition object counting and many more. For such huge scope of application image translation from adverse illumination condition to standard condition still faces challenges regarding detection in illumination changes such as low light specially during night time , motion blur, bad weather and in general real time detection. Despite the remarkably efficient and reliable image translation during standard conditions such as daytime, image translation at nighttime still faces challenge due to low contrast against the background hence affecting accuracy of the detected object and overall system performance. Another reason behind the efficient image translation not being successful is the lack of sufficient annotated dataset of nighttime. This research work aims to further explore the existing solution for the loss in efficiency for night time image translation using GAN.

1.1 Motivation

In this world humans are using highly impressive photo transfer application, we can feel face recognition, text to image translation, more realistic cartoon characters, photo to image and image to photo, three D printing, damaged photo recovery, automatic driving and many others. All such application involves computer vision related techniques and hence image to image transfer and augmentation are of highly use. Despite being highly useful for applications like advanced driving, object detection from night time image, virtual travelling experience during night time as well, night to day image translation faces still challenge due to the bad illuminating condition.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:03/March-2022 Impact Factor- 6.752 www.irjmets.com

Improving the application scope and to further expose this field seems to be a significant contribution for better development. Hence testing and deploying cycle GAN may be fruitful.

II. RELATED WORK

In [1], researchers proposes night to day image translation for virtual day driving experience during night time. They proposed daydriex image processing pipeline with GAN to achieve such virtual driving experience. They compare their result using FID and user perception scores. The obscured image in nighttime is augmented using position matching for specified location and merging day image for the same position. The steps they implemented are hint image retrieval, position matching, perspective matching, and image merging and image translation. They compared their proposed models output with cycleGANs output and found higher user perception score for their outcomes additionally they compared their output with cycleGAN using FID score. Lower the FID score higher success in image translation. In [2], authors attempts to review the solution for image translation problems faced during adverse condition such as night to day translation, varying illuminating conditions and bad weather image translations. Mostly all the translations related reviews are based on GAN models and its variants. They categorize the GAN and its variants and compare their pros and cons along with limitations. Dissertation by Kim [3] proposes semantic object conversion model using the change of local semantic objects categories at night to achieve better image based localization at night. Pix2pix and ToDayGAN method are compared in the research. They trained mask R-CNN for instance segmentation of night and day images and pix2pix GAN model for category wise night to day image conversion. For tackling the visual illumination change problem they used visual localization technique to obtain better performance. Instance segmentation and image conversion are the steps utilized during the night time image conversion process in [3]. Authors in [4] proposes novel ForkGAN model with introduction of fork branch in each generator stage to boost the performance of localization consequently the image enhancement in adverse illumination condition. They evaluated the output using FID and vision task metrics and claims to be very effective for object detection and image translation for adverse light condition. Researchers in[5] states that due to the environment of illumination lack, uneven lightening and artificial lightening, most of the existing image enhancement approaches fails to work well on nighttime images. To solve above problem authors in [5] has proposed GAN based framework for nighttime image enhancement. They present fusion network in which the dark channel prior based illumination compensation is employed. They combined several loss functions including the perpetual loss from the pre-trained VGG network [5]. In [6] author focuses on day to night image transfer for effective vehicle detection for night time driving and autonomous driving using AI. They propose AuGAN, a structure aware unpaired image to image translation network. This model detects by translating existing labeled data from its original domain to other ones. They applied both one stage YOLO and two stage faster R-CNN detectors in assessing how well the day to night transformation is done by each GAN Models in terms of vehicle detection. They use KITTI and SYNTHIA dataset for on road night time vehicle detection [6]. On [7] authors explores various image to image transformation method to transfer image from source domain to a target domain while preserving the content representation. They have analyzed key techniques of the existing image to image works and clarify the main progress to the community has made. They list out almost all evaluation metrics. Low light image enhancement using Generative adversarial network are used in [8]. Single image contrast enhancement SICE dataset are used in the research and additionally various datasets are clubbed together to create a paired training set for the research [8]. Encoder, decoder, discriminator, convolution 2D, Batch normalization and LeakyReLU are main features implemented in the proposed work. They used mean absolute error MAE as loss function for generator and Binary cross entropy for the discriminator network [8]. For solving the problem of place recognigition, localization and image translation for real life application like autonomous driving and robotic applications for adverse condition, researchers in reference [9] proposes ToDayGAN model implemented on Oxford RobotCar Dataset.

III. RESEARCH METHODOLOGY

3.1 Generative adversarial network GAN

Based on conditional generative adversarial network, cycle Generative adversarial network is general approach for unpaired image to image translation. Here with the two dataset, night time image and day time image, cycle Generative Adversarial Network is used to translate the image from night time to day time and day time to night

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:03/March-2022 Impact Factor- 6.752 www.irjmets.com

time. It is based on the conditional generative adversarial network, where a target image is generated, conditional on a given input image [10]. In this case, the cycle GAN changes the loss function so that the generated image is both reasonable in the content of the target domain, and is a reasonable translation of the input image [10]. Two discriminator-generator pair G_{ND}, D_{day} and G_{DN}, D_{night} are built to transfer night image to day and day to night. The generator discriminator pair are aimed to map image from night to day and day to night in cyclic manner. For generator G_{ND} and its corresponding discriminator D_{day} the adversarial loss is given by the formula below [11]

$$L_{Dadv}(G_{ND}, D_{day}, p_x, p_y) = E_{y\sim p_y}[(D_{day}(y) - 1)\ ] + E_{x\sim p(z)}[(D_{day}(G_{ND}(x))\ )] \quad (1)$$

where p_x is the sample distribution in night and p_y is the sample distribution in day [11].

3.1.1 Adversarial Loss

For generator G_{DN} and its associated discriminator D_{night} the adversarial loss is given by the equation below [11]

$$L_{Dadv}(G_{DN}, D_{night}, p_y, p_z) = E_{y\sim p_y}[(D_{night}(y) - 1)\ ] + E_{x\sim p(z)}[(D_{night}(G_{DN}(y))\ )] \quad (2)$$

3.1.2 Cycle Consistency loss

The cycle Consistency loss is given by [11]

$$L_{cyc}(G_{ND}, G_{DN}) = E_{x\sim p_x}[||G_{DN}(G_{ND}(x) - 1||) + E_{y\sim p_y}[||G_{ND}(G_{DN}(y) - 1||) \quad (3)$$

3.1.3 Identity Loss

Generator being multiclass convolutional neural network, identity loss forces network generator to use identity matrix on samples from source domain to destination domain (Night to Day). The identity loss calculated in similary manner to [11] is given by equation below

$$L_{idt}(G_{ND}, G_{DN}, p_x, p_y) = E_{x\sim p_x}[||G_{DN}(x) - x||] + E_{y\sim p_y}[||G_{ND}(y) - y||)] \quad (4)$$

Based on the experimental result, authors in [11] claims that the model may generate unreal result without L_{idt}.

L_{idt} is used to preserve the color composition between the input and output [11]. Now the total loss is given by

$$L_{total} = L_{Dadv}(G_{ND}, D_{day}, p_x, p_y) + L_{Nadv}(G_{DN}, D_{night}, p_y, p_z) + L_{cyc}(G_{ND}, G_{DN}) + L_{idt}(G_{ND}, G_{DN}, p_x, p_y) \quad (5)$$

3.2 Data-set

The night time and day time images taken from front camera of car on a road are archived and are available as open source dataset in kaggle. The dataset contains 14607 day time road images and 16960 night time road image. The images are RGB images with 640 pixel width and 260 in height. Both the horizontal and vertical resolution is 96 dpi and has bit depth of 24, meaning each image contains 8 bits per pixel for each of 3 channel.

3.3 Cycle Generative Adversarial Network for Night to Day and Day to night image transfer

Transferring Night images to day images using cycle Generative Adversarial Network requires pairs of generator and discriminator. Lets assume Generator G_{ND} maps Night image to Day image and and G_{DN} maps corresponding Generated Day image to Night image again as shown in block diagram below. D_{night} encourages G_{DN} to translate Day images into equivalent night images indistinguishable from night domain and D_{day} encourages G_{DN} to translate day images to night images indistinguishably from real night domain. Two adversarial discriminator D_{day} and D_{night} are needed. This process can further be elaborated using the diagram below.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:03/March-2022 Impact Factor- 6.752 www.irjmets.com

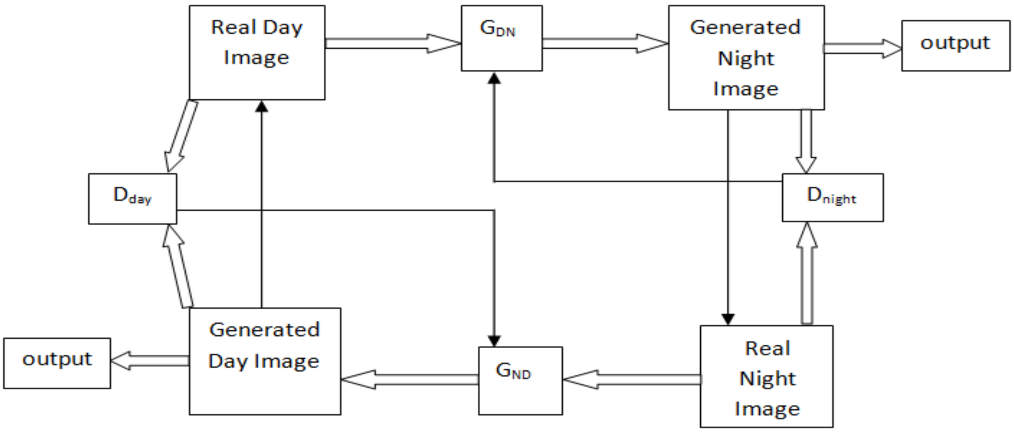

Figure 1: Block Diagram

To make it more simpler lets take night to day image transfer cycle. For night to day image transfer cycle to complete, night image are fetched as input to U-net Generator instead of random noise, the generator G_{ND} generates corresponding generated day image. Thus generated day image is fetched to patchGAN discriminator Dday, which is trained to distinguish between such generated and real day image untill the generator is good enough to fool the discriminator.

3.4 Data Pre-processing

The day time road images and night time road images are are randomly split into train and test images set for both day time and night time. All the images of size 640x260 are re-scaled to 256x256. Initially the images are cropped to 256x256x3 and then flipped. Geometrical transformation is performed to achieve better feature extraction in later stages.

3.5 U-net Generator

Basically U-net is a encoder decoder architecture structure with skip connection. Taking a real image to obtain segmentation mask or labels on every single pixel here U-net with skip connection is deployed. U-net is good at taking in an input image and mapping it to an output image. U-net generator takes the entire night image instead of noise vector like in the case of traditional GAN. The encoder decoder network first encode the multidimensional features. The encoder network then classifies the pixel and outputs the class. The middle bottleneck layer is used to attach the previously classified result to the image. During this process all the important information in the image is compressed from bottleneck and important features that will concatenated to the decoder. U-net uses skip connection from encoder to decoder. The skip connection from encoder to decoder skips the procedural stepwise flow of information so that the important features are fed directly to the output. In this research the encoder takes in an image of size 256x256 width by height with three channel RGB color. Thus segmented image goes through eight encoder blocks to compress that input and then the spatial sizes are down-sampled by half so that the output size of encoder becomes 128, 64 and so on divided by two in every encoder block. Each of these encoder blocks contains a convolution layer, a BN norm layer and Leaky ReLU activation. The convolution down-samples the night image with height and width by stride of 2. The decoder side contains eight decoder blocks with single stride. Eight blocks are required to produce the same size image since our encoder contains same numbers of blocks. The decoder is built up of transposed convolution followed by batch norm and Relu activation function. U-net integrates the distribution information from encoder into decoder using concatenation at each blocks. In summary U-net performs encoding decoding features that are mapped to each corresponding transposed form from encoder to decoder. The down-sampling encoder network in U-net network is fed with input image of size 256x256 which is preprocessed from the night time images dataset. The downsampling network is composed of eight convolution network with instance normalization and Relu activation function. The upsampling network consist of similar eight convolutional network. Instance normalization and relu activation function are used. figure 3 displays the Structural Diagram for Unet Generator. The side arrows represents the network to be skip connected. Basic operations during the image generation are convolution, normalization, activation, up-sampling, dropout and concatenation

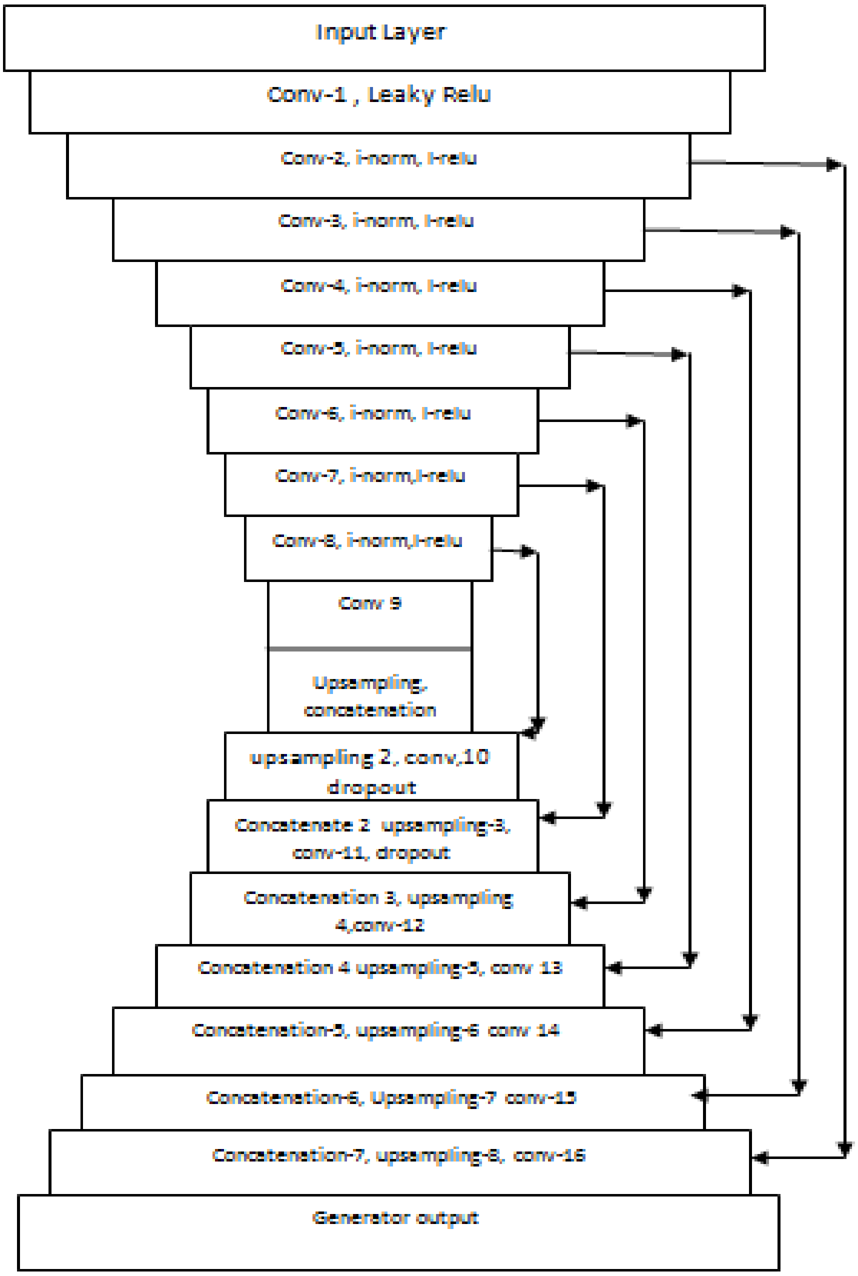

Figure 2: UNET Structural Diagram

3.6 Discriminator

The out put images generated from generator and the real image for corresponding domain are taken as input for the discriminator network. Being faster, capable of working on large data set and fewer parameters to fine tune the output PatchGAN has out weighted other discriminator network in image conversion. PatchGAN splits input image to small small patches so that discriminator can run convolutionally for all patches. The responses obtained are averaged to produce final outputs so that it can conclude wheather the input image is real or fake. Instead of signaling the output interms of 0 and 1, the PatchGAN outputs array to acknowledge the originality of the input image for respective day time image or night time image. The PatchGAN network takes real image as input with 256x256 dimension. There are five conovlution layer with instance normalization and leaky Relu Activation function.

3.6.1 PatchGAN

PatchGAN contains two dimensional convolution layers. In the first layer which is followed by leaky relu and maxpooling. similarly the second Convolution2D is followed by instance normalization. Leaky Relu and maxpooling is used upto 4th convolution. After taking images of size 257 by 256 size from real and generated dataset that image is patched into 70x70x3 sizes patches. The network structure diagram is shown in figure bleow. As Standard Convolution batch normalization Relu-Blocks are used by both generator and discriminator models in cycle GAN. [10].

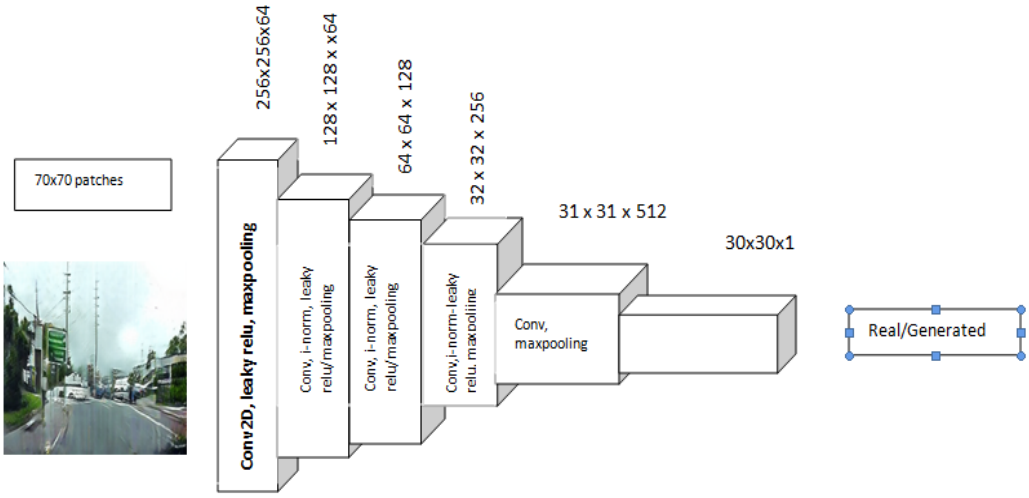

Figure 3: PatchGAN Structural Diagram

IV. RESULT AND DISCUSSION

The results obtained from model during writing this proposal are shown as below. The findings are presented here along with evaluation matrixs and outputs. The transformation performed helped to avoid over-fitting, because from experiment it is observed from the above loss plots that: from few epochs to larger epochs the generator loss and discriminator loss are constant. Fine tuned models output are shown as figure below. figure below shows some real vs generated image pairs upto 8th epoch.

Figure 4: Real vs generated image pairs in 8th epoch for local images

From above figures it is clear that due to the poor illumination on local images the result is slightly lower while with better illumination and image quality the generated images are near equivalent to the corresponding day images.

V. EVALUATION

The performance of the model is being evaluated with Frechet Inception Distance (FID), which is most common evaluation parameter for performance of GAN models and Structural Similarity SSIM Index and Mean square error MSE. The Mean square error is calculated between each pixels for the two images and SSIM calculated pixel densities. Since Mean square error generally tends to have higher arbitrary values it is not best suited for GAN evaluation.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-k(i,j)]2 \qquad (6)$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \qquad (7)$$

Structural Similarity (SSIM) index, helps us to compare the similarity in structural and features and is used to compare the similarity between generated and original image based on the perception. Figure below shows the SSIM for generated image obtained from 8 th epoch with datasize 16000 batch size 1 and learning rate 0.0002. The SSIM is found to be 0.84 for training set and 0.78 for testing image and its reconstructed image from generator. This test image is taken from out of training samples.

Figure 5: SSIM Score for real vs Generated image pairs

Another matrix being used for evaluation of this day night generative adversarial network is Frechet Inception Distance. The FID score is given by the formula

$$FID = d(X,Y) = (\mu_x - \mu_y)^2 + (\sigma_x - \sigma_y)^2 \qquad (8)$$

FID measures feature wise mean of real and generated image pairs distance. Lower the FID the generated image is more closer to the correcponding real image. The FID achieved so far is 265, from higher values signigying that model is being improved but it is still insufficient.

VI. CONCLUSION

Cycle Generative Adversarial Network using unet convolutional neural network as generator and PatchGAN convolutional neural as discriminator for data size 650 and for 16000 night time road images and day time road images are trained. With learning rate 0.0002, 0.0004, 0.0001 the model is trained for upto 92th epoch for small datasize and upto 22th epoch for large datasize as well. Adversarial loss, Cycle Consistency loss and Identity loss are evaluated. The training loss plot obtained and the output images as well are found to be plausible around 8 th epoch for datasize 16000 with batch size 1 and learning rate 0.0002. Evaluation of the generated and reconstructed image with Frechet Inception Distance and Structural similarity matrix is performed. The structural similarity Matrix best obtained for real input night image and reconstructed image during epoch 8 is 0.78, for image other than training datasets. Frechet inception distance obtained is 265. Equivalent day image generated for given night image shows the clear indication that the light in night image clearly affects the generated images accuracy. As Future work this model can be improved to well perform on images with modified generator and discriminator, and as well for more adverse lighting conditions. Image enhancement techniques before and after generation can be more plausible for practical applications.

VII. REFERENCES

[1] Euihyeok Lee and Seungwoo Kang. Daydriex: Translating nighttime scenes towards daytime driving experience at night. Applied Sciences, 11(5):2013, 2021.

[2] Aziz Alotaibi. Deep generative adversarial networks for image-to-image translation: A review. Symmetry, 12(10):1705, 2020.

[3] Dongyoun Kim, Sangwon Shin, Jinwoo Park, and Sung Shin. Development of a semantic scene conversion model for image-based localization at night. In Proceedings of the Conference on Research in Adaptive and Convergent Systems, pages 107–112, 2019.

[4] Ziqiang Zheng, Yang Wu, Xinran Han, and Jianbo Shi. Forkgan: Seeing into the rainy night. In Computer Vision– ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, pages 155–170. Springer, 2020.

[5] Yingying Meng, Deqiang Kong, Zhenfeng Zhu, and Yao Zhao. From night to day: Gans based low quality image enhancement. Neural Processing Letters, 50(1):799–814, 2019.

[6] Che-Tsung Lin, Sheng-Wei Huang, Yen-Yi Wu, and Shang-Hong Lai. Gan-based day-to-night image style transfer for nighttime vehicle detection. IEEE Transactions on Intelligent Transportation Systems, 22(2):951–963, 2020.

[7] Yingxue Pang, Jianxin Lin, Tao Qin, and Zhibo Chen. Image-to-image translation: Methods and applications. arXiv preprint arXiv:2101.08629, 2021.

[8] Lloyd Dsouza3 Kunal Meher4 Daryl Fernandes1, Joel Dsouza2. Low-light image enhancement using generative adversarial networks. IRJET,2395-0056, 2021.

[9] Asha Anoosheh, Torsten Sattler, Radu Timofte, Marc Pollefeys, and Luc Van Gool. Night-to-day image translation for retrieval-based localization. In 2019 International Conference on Robotics and Automation (ICRA), pages 5958–5964. IEEE, 2019.

[10] Jason Brownlee. Generative adversarial networks with python: deep learning generative models for image synthesis and image translation. Machine Learning Mastery, 2019.

[11] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 994–1003, 2018.