DNIT: Enhancing Day-Night Image-to-Image Translation through Fine-Grained Feature Handling (Student Abstract)

Hanyue Liu[1], Haonan Cheng[2*], Long Ye[2,3]

[1]School of Information and Communication Engineering, Communication University of China, Beijing, China
[2]State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China
[3]School of Data Science and Media Intelligence, Communication University of China, Beijing, China
{hanyueliu, haonancheng, yelong}@cuc.edu.cn

Abstract

Existing image-to-image translation methods perform less satisfactorily in the "day-night" domain due to insufficient scene feature study. To address this problem, we propose DNIT, which performs fine-grained handling of features by a nighttime image preprocessing (NIP) module and an edge fusion detection (EFD) module. The NIP module enhances brightness while minimizing noise, facilitating the extraction of content and style features. Meanwhile, the EFD module utilizes two types of edge images as additional constraints to optimize the generator. Experimental results show that we can generate more realistic and higher-quality images compared to other methods, proving the effectiveness of our DNIT.

Introduction

Image-to-image translation (I2IT) converts an input image from a source domain to a target domain. Current I2IT methods, like CycleGAN (Zhu et al. 2017), excel in general-purpose tasks but exhibit translation errors when dealing with the specific "day-night" scenario. In this case, some methods utilize additional information to improve translation performance. For example, SPN2D-GAN (Li and Guo 2022) proposes a semantic prior generator that utilizes semantic information as weak supervision. However, there are still some localized problems of mistranslations. This is mainly because the methods do not fully consider the characteristics of the image domain and fail to improve for specific scenes. In "day-night" scenes, nighttime images have limited visibility due to low illumination, making obtaining features difficult. In addition, the scene's complexity increases the risk of losing the structural details of the image.

To address these challenges, we propose DNIT, which focuses on fine-grained feature handling by designing nighttime image preprocessing (NIP) and edge fusion detection (EFD) modules. Our DNIT significantly improves the visibility of nighttime images and effectively utilizes edge information to enhance the quality of I2IT processing.

Our Method

Our DNIT architecture is shown in Figure 1, with the NIP and EFD module details as follows.

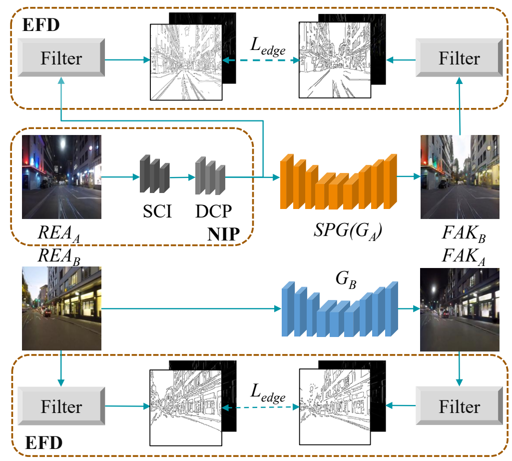

Figure 1: The architecture of our DNIT.

NIP Module To better learn accurate mapping relationships from low-light nighttime images, we designed the NIP module. It comprises a self-calibrated illumination (SCI) learning network and a dark channel prior (DCP) defogging network. SCI is a weight-sharing cascaded light learning network and utilizes a self-calibrated module for acceleration. DCP utilizes statistical principles to calculate the dark channel, estimate the light component, calculate the transmittance matrix, and accomplish defogging. Compared to single low-light enhancement technologies, our NIP module is more versatile and effective in image processing.

In the initial stage of I2IT, we preprocess the original nighttime image REA_A by applying SCI to enhance brightness. Next, we use DCP to eliminate the additional noise introduced during brightening. Finally, we feed the preprocessed image into the semantic prior generator $(SPG(G_A))$ for conversion. Our NIP module enhances image brightness, reduces additional noise, and generates high-quality preprocessed images. This process allows for more efficient exploration and extraction of information from dark areas, providing the generator with enhanced content and style features to improve I2IT performance.

EFD Module To reduce the loss of edge details during the I2IT process, we propose the EFD module that utilizes mul-

*Corresponding author

Model	FID↓	KID↓	BRISQUE↓	NIQE↓
CycleGAN	65.77	2.82	31.65	4.30
CUT	87.68	4.63	33.08	4.29
NICE-GAN	83.21	4.44	34.05	4.33
DCLGAN	80.73	4.28	31.30	4.45
ToDayGAN	69.60	3.06	29.09	3.95
SPN2D-GAN	65.54	2.46	20.36	**2.58**
Ours	**64.72**	**2.23**	**19.04**	2.65

Table 1: Quantitative comparison of results, the lower the better for all metrics.

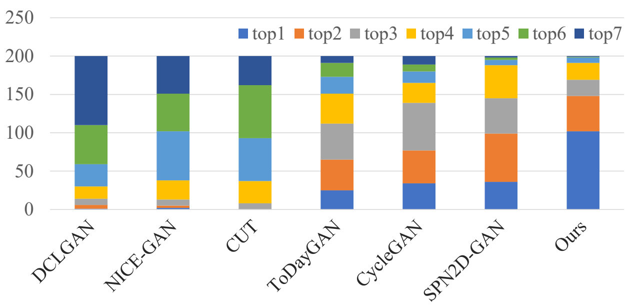

Figure 2: User study results. Top 1 is the best quality.

tiple edge information constraint generators. We design the EFD module combining Sobel and Canny filters for detecting the edge maps of real and generated images and using L1 loss to compute the distance between the edge maps. The specific edge loss is as follows:

$$\mathcal{L}_{\text{edge}} = \parallel S_x - S_{G_A(x)} \parallel_1 + \parallel S_y - S_{G_B(y)} \parallel_1 + \parallel C_x - C_{G_A(x)} \parallel_1 + \parallel C_y - C_{G_B(y)} \parallel_1, \quad (1)$$

where x and y are real images, $G_A(x)$ and $G_B(y)$ are generated images, $S(\cdot)$ is the Sobel edge detection images and $C(\cdot)$ is the Canny edge detection images. By using the Sobel and Canny filters jointly, we can exploit their potential complementarity and thus improve the continuity and accuracy of the edge images. Adopting the EFD module can effectively enhance the supervision of the I2IT process and facilitate the generation of more realistic images.

Experiments and Results

We compare DNIT with previous I2IT methods on the ACDC dataset, including CycleGAN, NICE-GAN (Chen et al. 2018), ToDayGAN (Anoosheh et al. 2019), CUT (Park et al. 2020), DCLGAN (Han et al. 2021) and SPN2D-GAN. As shown in Table 1, compared with the baseline SPN2D-GAN, our DNIT reduces FID from 65.54 to 64.72, KID from 2.46 to 2.23, and BRISQUE from 20.36 to 19.04 while keeping the number of model parameters unchanged. The NIQE is slightly higher, which indicates that the quality of the natural scene statistics (NSS) feature may be lower. To further verify the naturalness, we conduct a user study with 20 evaluators who rank the quality of 70 images. Unlike metrics focusing on aspects such as similarity and statistical properties, this user study provides a comprehensive assessment based on visual perception. As shown in Figure 2, our method ranks first at 51%, proving that we can generate better-quality images. Figure 3 shows the results of the qualitative comparison. Our method can improve the problems of texture error, structure loss, and color deviation in previous methods. Furthermore, our DNIT exhibits effective day-to-night image translation capabilities, similar to the night-to-day direction (except for the NIP module and $SPG(G_A)$).

Figure 3: Results of the comparison experiments.

Conclusions and Future Work

Our DNIT enhances the “day-night” image translation, highlighting the efficacy of fine-grained feature handling. In future work, we aim to improve the adaptive Canny threshold selection for better generalization across diverse scenes.

Acknowledgments

This work is supported by the Natural Science Foundation of China under grant No. 62201524, No. 61971383, and No. 62271455.

References

Anoosheh, A.; Sattler, T.; Timofte, R.; Pollefeys, M.; and Van Gool, L. 2019. Night-to-day image translation for retrieval-based localization. In *ICRA*, 5958–5964.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 801–818.

Han, J.; Shoeiby, M.; Petersson, L.; and Armin, M. A. 2021. Dual contrastive learning for unsupervised image-to-image translation. In *CVPR Workshops*, 746–755.

Li, X.; and Guo, X. 2022. SPN2D-GAN: Semantic Prior Based Night-to-Day Image-to-Image Translation. *TMM*.

Park, T.; Efros, A. A.; Zhang, R.; and Zhu, J.-Y. 2020. Contrastive learning for unpaired image-to-image translation. In *ECCV*, 319–345. Springer.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2223–2232.