# N2D-GAN: A NIGHT-TO-DAY IMAGE-TO-IMAGE TRANSLATOR

*Xiaopeng Li, Xiaojie Guo, Jiawan Zhang**

College of Intelligence and Computing, Tianjin University, Tianjin, China
{lxpalliance, jwzhang}@tju.edu.cn, xj.max.guo@gmail.com

2022 IEEE International Conference on Multimedia and Expo (ICME) | 978-1-6654-8563-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICME52920.2022.9859906

## ABSTRACT

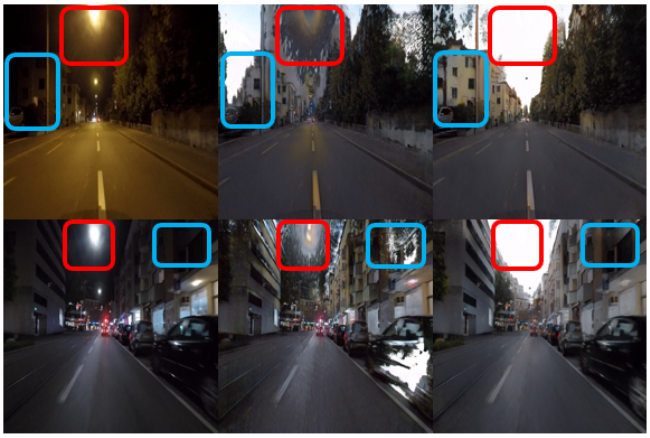
Existing image-to-image translation methods can effectively deal with simple scenes or styles, such as horses-to-zebra, cat-to-dog, and summer-to-winter. However, the performance of night-to-day (N2D) translation remains unsatisfied due to imbalanced/poor visibility and thus translation ambiguity, although some progress has been made by GAN-based methods recently. This paper proposes a CycleGAN-based N2D translation scheme, namely N2D-GAN, in a weakly-supervised manner with consideration of both image and semantic information. Specifically, we first adjust the brightness of nighttime images to boost the visibility, so that the generator can better extract content information. Then, the generator processes the translation by enforcing results to follow the distribution of daytime images in both image and semantic domains. Besides, the cycle consistency is introduced to preserve the fidelity between translations from two directions. Experimental results demonstrate that our strategy outperforms other state-of-the-art N2D methods both quantitatively and qualitatively.

***Index Terms*—** Image-to-image/Night-to-day translation, generative adversarial network, weakly-supervised learning

## 1. INTRODUCTION

The goal of image-to-image translation (I2IT) is to learn a mapping from a source domain to a target domain, e.g., summer to winter, and image to semantic label, which has shown its wide applicable scenarios, including style transfer, semantic label generation, super-resolution, domain adaptation [1], image enhancement [2], and data augmentation. Over past years, a series of I2IT methods in supervised settings (using paired data) have shown impressive performance. However, collecting a large number of paired pixel-to-pixel data is difficult and expensive.

In order to achieve I2IT in the absence of paired examples, unsupervised methods have been developed, the models of which can be trained using unpaired data from the source and target domains. Recently, unpaired I2IT methods have made great progress in many style transfer tasks, e.g., summer to winter, day to night, sunny to rainy. However, these methods only perform well on natural landscapes and single-category images with less detailed and structural information. From the examples in Fig. 1, we can see that obvious visual artifacts left in complex urban scene images when applying to the night-to-day (N2D) translation.



Input CycleGAN Ours

**Fig. 1**. From left to right: two real-world nighttime images, the translation results by CycleGAN and our proposed N2D-GAN method, respectively. Over-exposed regions are highlighted in the red boxes, while under-exposed regions are highlighted in the blue boxes.

The main reason why these existing unpaired I2IT methods cannot achieve pleasant performance for N2D cases comes from the poor visibility of nighttime images, leading to translation ambiguity. Different from other images, nighttime real-world images are of uneven intensity distribution. In other words, those over-exposed and under-exposed regions appear at different spatial positions in nighttime images, which make content extraction more challenging. Also, noise can also disturb the image-to-image translation. Additionally, the high-level semantic information of nighttime images is inevitably destroyed by the noise and low-light conditions. Therefore, the daytime images generated by the existing unpaired image-to-image translation methods would contain massive texture errors and lack of structural information, resulting in low-quality translation results.

To alleviate the aforementioned issue, we propose a novel

*corresponding authors.

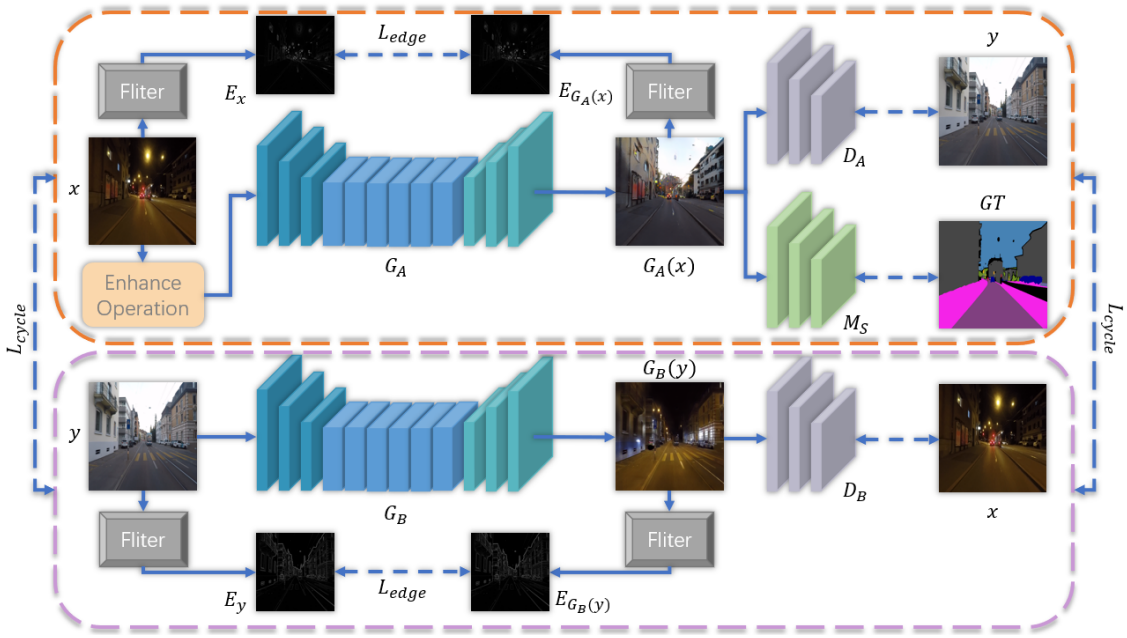



**Fig. 2**. Architecture of our proposed N2D-GAN, which includes two generators $G_A$, $G_B$, two discriminators $D_A$, $D_B$, and a semantic segmentation module $M_S$.

CycleGAN-based night-to-day translation method, termed as N2D-GAN. In a weakly-supervised fashion, our method can effectively improve the performance of N2D translation. To achieve the goal, We employ the traditional low-light enhancement method to enhance the brightness of under-exposed regions in order to extract richer structure and texture information. Aiming to make sufficient use of the detailed information within the original image, we calculate the edge loss between the gradient maps of input and output to guarantee the generated image retaining key edge messages. Since the ACDC dataset [3] provides semantic labels for nighttime images, we can bring additional high-level semantic label to the task for better regularize the generator. Furthermore, we impose constraints on the texture and structural feature-level information between the generated daytime image and the original nighttime image to make generated images look more realistic with well preserved content.

Concretely, our contributions can be summarized as:

- We propose a novel weakly-supervised I2IT network, called N2D-GAN, which significantly boost the N2D performance by mitigating the negative effect from poor visibility (ambiguity) of nighttime images.
- We use the gradient map of the original nighttime image to constrain the generated daytime image, which can help to preserve more details, and thus make the results more realistic.
- Besides the constraint in image domain, we adopt the high-level semantic segmentation information to provide additional guidance to the translator, which effectively reduce the semantic chaos in translated images.
- Extensive experiments are conducted to demonstrate the efficacy of our design, and reveal its superiority over other competitors on the N2D task both quantitatively and qualitatively.

## 2. RELATED WORK

Due to limited space, this section will briefly review representative GAN-based and unpaired I2IT approaches, which are closely related to this work.

**Generative adversarial networks (GANs).** The framework of GANs [4] framework can be divided into two sub-networks, a generator and a discriminator, which are trained together through the two-player minimax game. The purpose of the generator is to generate realistic images that can confuse the discriminator, while the purpose of the discriminator is to distinguish whether the input image is a real image or a generated image. A series of GANs have been proposed to generate more realistic images. DCGAN [5] uses the convolutional structure in the discriminator and generator to make the network training more stable. WGAN [6] concludes that Wasserstein distance is nicer than Jensen-Shannon divergence. Conditional GAN [7] injects extra information to guide the image generation process.

**Unpaired image-to-image translation.** The unpaired image-to-image translation aims to learn the bidirectional mapping function between the source domain and target domain without the requirement of paired training data. Specifically, CycleGAN [8] can effectively learn the mapping between two domains with the help of the proposed cycle consistency loss. MUNIT [9] recombines the content code of im-

age with a random style code sampled from the style space of the target domain to translate an image to another domain. SCAN [10] can enable higher resolution image-to-image translation in a coarse-to-fine fashion. U-GAT-IT [11] proposes a novel method for unsupervised image-to-image translation with a new attention module and a new normalization function AdaLIN. AttentionGAN [12] proposes to use the attention mechanism for unpaired image-to-image translation. CUT [13] and DCLGAN [14] can achieve impressive performance based on contrastive learning. NICE-GAN [15] designs a more compact and more effective architecture by reusing the discriminator for encoding.

Unfortunately, most of existing unpaired image-to-image translation methods have numerous as well as evident visual errors and incorrect textures in night-to-day translation results. By using additional semantic and gradient information, our weakly-supervised learning image-to-image translation approach can make it better to reduce the appearance of visual errors and artifacts.

## 3. METHOD

### 3.1. Model Architecture

The unpaired I2IT aims to learn the mapping between domain A and domain B. Specifically, we set the night domain to domain A and the day domain to domain B during the night-to-day translation task. Similar to the CycleGAN [8] architecture, N2D-GAN mainly includes two generators $G_A$, $G_B$, two discriminators $D_A$, $D_B$, and a semantic segmentation module $M_S$. The generator $G_A$ learns a mapping $f_{A\to B}$ from the night domain A to the day domain B, while the generator $G_B$ learns a mapping $f_{B\to A}$ from the day domain B to the night domain A. The discriminator $D_A$ and $D_B$ aim to distinguish the generated image from the generator and the real image from the dataset. Different from CycleGAN [8], we performed an image enhancement operation on the input image before sending it to the generator $G_A$, which allows the generator to capture more structure and texture information. Besides, we add the filter to extract the edge map of the original image $x$, $y$, and the generated image $G_A(x)$, $G_B(y)$ respectively. Furthermore, aiming at retaining the semantic information of the night-to-day transition effectively, we utilize a semantic segmentation module $M_S$ to combine the high-level semantic segmentation task with the image translation task. The entire N2D-GAN framework is shown in Fig. 2.

### 3.2. Semantic Segmentation Module

Real-world nighttime images lose high-level semantic information due to the existence of under-exposed regions and over-exposed regions. Because the neural network cannot figure out what objects are hidden behind these over-exposed/under-exposed regions, incorrect textures and artifacts are generated, e.g., black artifacts around street lights, trees that appear above buildings, buildings with sky texture and so on.

To alleviate the aforementioned issue, we propose a semantic segmentation module to impose the semantic constraint on the generator. The image generated by the generator $G_A$ will be extracted the semantic map by the semantic segmentation module, and compared with the semantic map of the original image. The weakly-supervised learning approach using high-level semantic information for additional supervision can prevent the semantics of the generated daytime image from being changed effectively. Specifically, we use the DeepLabv3+ [16] model pre-trained on the Cityscapes dataset as the semantic segmentation module $M_S$.

### 3.3. Loss Function

The loss function of our N2D-GAN mainly consists of four components, i.e. edge loss, feature loss, semantic loss, and adversarial loss. In the following, we will introduce these loss terms respectively.

**Edge loss.** Extensive dark regions exist in the real-world nighttime image, which makes it difficult for the generator to extract image details. Inspired by the edge detection, we noticed that the edge information of the image can help to improve the quality of the generated image on our night-to-day image translation. Therefore, we propose to leverage the edge information of both the original image and the generated image as additional supervision to train the translator. We use the L1 loss to calculate the distance between the edge maps of both the original image and the generated image. The specific edge loss function formula can be listed as follows:

$$\mathcal{L}_{edge} = \left\|E_x - E_{G_A(x)}\right\|_1 + \left\|E_y - E_{G_B(y)}\right\|_1, \quad (1)$$

where $E_x$ and $E_y$ are edge maps extracted from original images, $E_{G_A(x)}$ and $E_{G_B(y)}$ are edge maps extracted from generated images. Moreover, because the edge information extracted directly from the nighttime image is less, we recommend performing an image enhancement operation on the nighttime image to enhance the illumination in advance. Considering the limited memory and the performance of different image enhancement approaches, we properly choose the gamma correction as the image enhancement operation.

**Feature loss.** From one point of view, an image can be decoupled into the structure/content component and the texture/style component. Based on the idea of decoupling content and style, numerous unsupervised domain adaptation semantic segmentation methods [17] and style transfer methods [18] have been proposed in recent years. Inspired by style transfer [18], we propose the feature loss, which imposes constraints on structure feature and texture feature respectively. The feature loss term can be expressed as follows:

$$\mathcal{L}_{feature} = \mathcal{L}_{perc}(x^s, \hat{x}^s) + \mathcal{L}_{perc}(x^t, \hat{x}^t), \quad (2)$$

where $x$ is the input image and $\hat{x}$ is the generated image. Concretely, the $\mathcal{L}_{perc}$ is defined as a weighted sum of L1 differences between feature representations extracted from a pre-trained VGG-19 network. Following the [17], the $\mathcal{L}_{perc}$ as defined in the following shape:

$$\mathcal{L}_{perc} = \sum_{l \in L} \frac{w^l}{N^l} \left\| \phi^l(x) - \phi^l(\hat{x}) \right\|_1, \tag{3}$$

where $\phi(\cdot)$ is the VGG network, $N$ is the number of pixels, $w$ is the weight of perceptual loss, and $l$ is a concrete layer of $L(\{relu1\_1, relu2\_1, relu3\_1, relu4\_1, relu5\_1\})$.

**Semantic loss.** The semantic segmentation module guides the generator $G_A$ to generate the daytime image which can retain the high-level semantic information of the original nighttime image through the semantic loss function. Specifically, we use the cross-entropy loss as our semantic loss function. The semantic loss can be written as follows:

$$\mathcal{L}_{seg} = -\sum_{i=1}^{n} y_i \log \widehat{y}_i. \tag{4}$$

**Adversarial loss.** Same as other generative adversarial networks, we use two discriminators $D_A$ and $D_B$ for adversarial learning. We adopt the following least-squares loss to enforce the generated images to be close to the real images.

$$\mathcal{L}_{adv} = (D_A(G_A(x)) - r)^2 + (D_B(G_B(y)) - r)^2. \tag{5}$$

The respective objective functions of the discriminators $D_A$ and $D_B$ are defined by ($r = 1$ and $f = 0$):

$$\min_{D_A} \mathcal{L}_{D_A} = \frac{1}{2}(D_A(x) - r)^2 + \frac{1}{2}(D_A(G_A(x)) - f)^2, \tag{6}$$

$$\min_{D_B} \mathcal{L}_{D_B} = \frac{1}{2}(D_B(y) - r)^2 + \frac{1}{2}(D_B(G_B(y)) - f)^2. \tag{7}$$

Inspired by the Cycle-GAN principle, N2D-GAN still needs the cycle consistency constraint and the identity constraint as the basic objective. Specifically, the cycle consistency loss given as follows:

$$\mathcal{L}_{cycle} = \|G_B(G_A(x)) - x\|_1 + \|G_A(G_B(y)) - y\|_1, \tag{8}$$

while the identity loss is defined as follows:

$$\mathcal{L}_{idt} = \|G_A(y) - y\|_1 + \|G_B(x) - x\|_1. \tag{9}$$

In summary, the whole training loss of the proposed N2D-GAN is defined as follows:

$$\mathcal{L}_{N2D-GAN} = \mathcal{L}_{GAN} + \lambda_e \mathcal{L}_{edge} + \mathcal{L}_{feature} + \mathcal{L}_{seg}, \tag{10}$$

where the $\lambda_e$ means the weight of the edge loss. The $\lambda_e$ is set to 10 in the experiments to keep all loss terms with the same magnitude. The $\mathcal{L}_{GAN}$ can be described as follows:

$$\mathcal{L}_{GAN} = \mathcal{L}_{adv} + \lambda_{idt} \mathcal{L}_{idt} + \mathcal{L}_{cycle} + \mathcal{L}_{D_A} + \mathcal{L}_{D_B}, \tag{11}$$

where the $\lambda_{idt}$ means the weight of the identity loss and is set to 5 by default.

# 4. EXPERIMENTS

## 4.1. Experimental Details

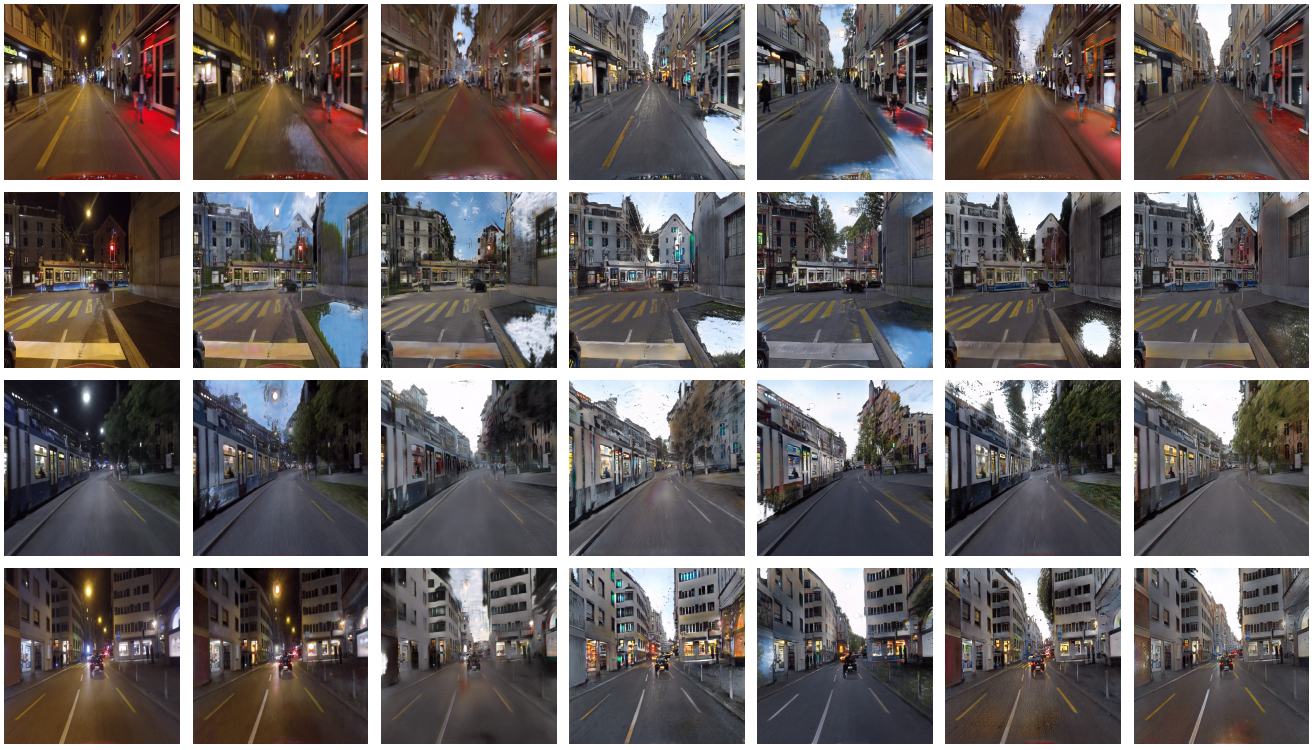
We implement the proposed N2D-GAN using PyTorch on a single Nvidia 2080Ti GPU. In the training phase, N2D-GAN is optimized by Adam in 200 epochs, the parameters of which are set as $\beta_1 = 0.5$, $\beta_2 = 0.999$, and the learning rate is set to 0.0002. Same as CycleGAN, after 100 epochs, the learning rate will linearly decrease to zero within 100 epochs. The crop size is set to $256 \times 256$ and the batch size is set to 1. The generator is an encoder-decoder structure with 9 residual blocks, while the discriminator is $70 \times 70$ PatchGANs, both with instance normalization. For the semantic segmentation module, we use the ResNet50-DeepLabv3+ network pre-trained on the Cityscapes dataset. All models in the experiments are trained with 506 nighttime and 506 daytime images from the training and validation set of ACDC-night dataset, while 500 nighttime and 500 daytime images from the test set of ACDC-night dataset are used to evaluate the quality of generated images.

## 4.2. Datasets and Evaluation Metrics

The ACDC dataset consists of a large set of 4006 images which are equally distributed between four common adverse conditions: fog, nighttime, rain, and snow. We choose the ACDC-night dataset as the main experimental dataset, which contains 1006 nighttime images, 506 semantic labels for nighttime images, and 1006 daytime images with a resolution of $1920 \times 1080$. The Cityscapes dataset is used to train our semantic segmentation network, which is a large-scale urban-scene dataset, holding high-quality pixel-level annotations of 5K images and 20K coarsely annotated images with a high resolution of $2048 \times 1024$. Following other I2IT methods, we employ the Fréchet Inception Distance (FID) as the evaluation metric to evaluate the quality of the images we generated. Lower FID is better, corresponding to generated images more similar to the real. We use the generated daytime images to train the semantic segmentation network and use the mean Intersection-over-Union (mIoU) metric to evaluate the prediction results to illustrate how much key semantic information is retained in the generated images.

## 4.3. Comparisons

For the purpose of fairly comparing with other I2IT methods, the experimental results of all methods were trained 200 epochs on the ACDC dataset based on the default settings. The values of metrics on the ACDC test set are exhibited in Table 1. We compare our N2D-GAN with some representative methods on unpaired I2IT task, including CycleGAN [8], NICE-GAN [15], DCLGAN [14], ToDayGAN [19], and CUT [13]. We can clearly see that most I2IT methods do not perform well on night-to-day translation. Obviously, our N2D-GAN achieves state-of-the-art performance on unpaired



(a) Input (b) CycleGAN [8] (c) NICE-GAN [15] (d) DCLGAN [14] (e) ToDayGAN [19] (f) CUT [13] (g) Ours

**Fig. 3**. Experimental results of night-to-day translation on ACDC-night test set.

Baseline $+M_S$ $+M_S + \mathcal{L}_f$ N2D-GAN

**Fig. 4**. Visual comparison between different variations of our N2D-GAN. $\mathcal{L}_f$ means the feature loss, and $M_S$ means the semantic segmentation module.

night-to-day translation task. To better illustrate the superiority of our method, we display some visual results of different methods in Fig. 3. The image translation of CycleGAN may fail on some complex scenes, resulting in images that are still in the night style. The images generated by DCLGAN and NICE-GAN are not ideal for preserving original image details. NICE-GAN, DCLGAN, ToDayGAN, and CUT can all generate obvious artifacts and incorrect textures. Compared with the above methods, our N2D-GAN generated images with fewer visual artifacts and higher image quality.

Experimental results of the extra semantic segmentation task are also shown in Table 1. We uniformly generate $960 \times 540$ daytime images as the training and validation set of the DeepLabv3+. We consistently use these images to train 40,000 iterations of the segmentation network to conduct a fair comparison. Although ToDayGAN achieves pleasant performance in the FID metric, the low mIoU value indicates that the generated images can not retain the pivotal semantic information of the original image well. It is worth mentioning that the prediction results of the semantic segmentation network trained by daytime images generated by our N2D-GAN can reach 38.52 mIoU, while the prediction results of other methods are all less than 35 mIoU. Compared with other methods, our N2D-GAN can not only generate more realistic daytime images, but also retain more vital semantic information.

### 4.4. Ablation Study

We conduct ablation studies on the ACDC-night test set in Table 2 to demonstrate the validity of the different components of our N2D-GAN. Taking CycleGAN as our baseline, we add our proposed edge loss, semantic segmentation module, and feature loss on it separately. The results show that any component of N2D-GAN can improve the quality of the generated daytime image. We noticed that the semantic segmentation module we proposed was most helpful in improving the quality of the generated image, which clearly demonstrates the importance of semantic information for image-to-image translation task. Moreover, we combine three components of N2D-GAN in pairs, and the experimental results reveal that any two of the components we proposed can help improve the quality of generated daytime images. Finally, the full N2D-GAN can achieve the best performance, with a 12.33 FID reduction compared to the baseline.

To better illustrate the effectiveness of our approach, we visualize and compare some different variations of our N2D-GAN, which is shown in Fig. 4. By leveraging the semantic constraint, the semantic segmentation module can signif-

**Table 1**. Quantitative comparison in FID and mIoU. For the FID, lower is better, while for the mIoU, higher is better.

| Model | FID↓ | mIoU↑ |
|---|---|---|
| CycleGAN [8] (ICCV'17) | 75.91 | 34.25 |
| DCLGAN [14] (CVPR'21) | 78.15 | 34.97 |
| CUT [13] (ECCV'20) | 75.58 | 34.81 |
| NICE-GAN [15] (CVPR'20) | 73.20 | 31.14 |
| ToDayGAN [19] (ICRA'19) | 67.56 | 32.57 |
| N2D-GAN | **63.58** | **38.52** |

**Table 2**. Ablation study on different configurations.

| Components | | | FID↓ |
|---|---|---|---|
| $\mathcal{L}_{edge}$ | $M_S$ | $\mathcal{L}_{feature}$ | |
| × | × | × | 75.91 |
| ✓ | × | × | 71.14 |
| × | ✓ | × | 67.85 |
| × | × | ✓ | 70.91 |
| ✓ | ✓ | × | 66.20 |
| ✓ | × | ✓ | 69.04 |
| × | ✓ | ✓ | 65.97 |
| ✓ | ✓ | ✓ | 63.58 |

icantly remove the widespread artifact caused by the over-exposed region in the sky. We can distinctly see that feature loss can further correct the wrong sky texture in the building. Besides, as can be seen from the regions highlighted by blue boxes in Fig. 4, adding edge loss can further remove some small artifacts with the help of edge information. Finally, a complete N2D-GAN can achieve the best visual performance.

## 5. CONCLUSION

In this paper, we have proposed an unpaired night-to-day translation method, namely N2D-GAN, which is a weakly-supervised learning method via exploiting extra semantic information and edge information. Semantic information is utilized by semantic segmentation module, while edge information is utilized by edge extraction filter. Additionally, our method uses the feature loss to make the image more realistic. Extensive ablation studies and detailed analysis have revealed the necessity and effectiveness of each component of our approach. Qualitative and quantitative results have demonstrated that our N2D-GAN can generate satisfactory results compared with other methods.

# Acknowledgement

This work was supported by the study and demonstration application of organization and service of cultural heritage knowledge graph, National Key Research and Development Program of China (Grant No. 2019YFC1521200), the National Natural Science Foundation of China under Grant no. 62072327, Grant no. 62172295, and TSTC under Grant no. 20JCQNJC01510.

## 6. REFERENCES

[1] J. Hoffman, E. Tzeng, T. Park, J.Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *ICML*, 2018.

[2] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE TIP*, vol. 30, pp. 2340–2349, 2021.

[3] C. Sakaridis, D. Dai, and L. Van Gool, "Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding," *arXiv:2104.13395*, 2021.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014.

[5] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv:1511.06434*, 2015.

[6] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *ICML*, 2017.

[7] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv:1411.1784*, 2014.

[8] J.Y. Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017.

[9] X. Huang, M.Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *ECCV*, 2018.

[10] M. Li, H. Huang, L. Ma, W. Liu, T. Zhang, and Y. Jiang, "Unsupervised image-to-image translation with stacked cycle-consistent adversarial networks," in *ECCV*, 2018.

[11] J. Kim, M. Kim, H. Kang, and K. Lee, "U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," *arXiv:1907.10830*, 2019.

[12] H. Tang, H. Liu, D. Xu, P. Torr, and N. Sebe, "Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks," *IEEE TNNLS*, 2021.

[13] T. Park, A. Efros, R. Zhang, and J.Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *ECCV*, 2020.

[14] J. Han, M. Shoeiby, L. Petersson, and M. Armin, "Dual contrastive learning for unsupervised image-to-image translation," in *CVPR*, 2021.

[15] R. Chen, W. Huang, B. Huang, F. Sun, and B. Fang, "Reusing discriminators for encoding: Towards unsupervised image-to-image translation," in *CVPR*, 2020.

[16] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018.

[17] W.L. Chang, H.P. Wang, W.H. Peng, and W.C. Chiu, "All about structure: Adapting structural information across domains for boosting semantic segmentation," in *CVPR*, 2019.

[18] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016.

[19] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, "Night-to-day image translation for retrieval-based localization," in *ICRA*, 2019.