

# Day-to-Night Road Scene Image Translation Using Semantic Segmentation

Seung Youp Baek<sup>ID</sup> and Sungkil Lee<sup>† ID</sup>

Sungkyunkwan University, Suwon, South Korea

## Abstract

We present a semi-automated framework that translates day-time domain road scene images to those for the night-time domain. Unlike recent studies based on the Generative Adversarial Networks (GANs), we avoid learning for the translation without random failures. Our framework uses semantic annotation to extract scene elements, perceives a scene structure/depth, and applies per-element translation. Experimental results demonstrate that our framework can synthesize higher-resolution results without artifacts in the translation.

## CCS Concepts

• Computing methodologies → Computational photography; Image processing;

## 1. Introduction

Image-to-image translation is a problem where the goal is to generate a new image based on the given image. Since the introduction of Generative Adversarial Networks (GANs), remarkable results have been demonstrated for many tasks, such as edges to photo, labels to facade, and day to night [IZZE17]. Day-to-night domain adaptation had not received much attention yet, but several GAN-based methods are proposed recently and improvements are on-going.

Despite the success of learning-based methods, it is often difficult to tune the result for the desired outcome as it is unpredictable. Also, datasets with desired properties can be hard to obtain, as they are not always publicly available. Moreover, the results can be erroneous when unknown scenes are fed as inputs.

In this poster, we present a semi-automatic framework for the day-to-night image translation of the road scenes. We design our framework to avoid the learning. Instead, we utilize the semantic annotation to extract scene elements in the image. Thereby, we get able to perceive the scene, and can perform per-element translation. First, we generate a night-time image with per-element brightness adjustments. Then, single image-based feature extractions are performed to enhance local contrast in lightings associated with the coarse depth. Here, with the aid of semantic annotation, features are only extracted from selected elements, unintended translations are likely to be avoided. Therefore, more adjustable and predictable results can be obtained by our framework, which indicate feasibility for a further application, such as night-time dataset generation.

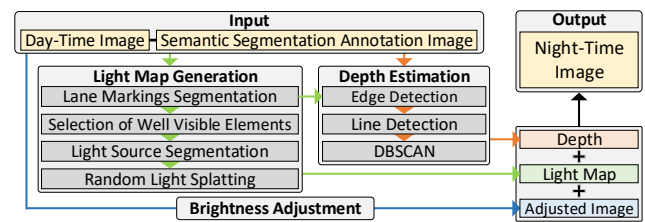


Figure 1: The overview of our framework.

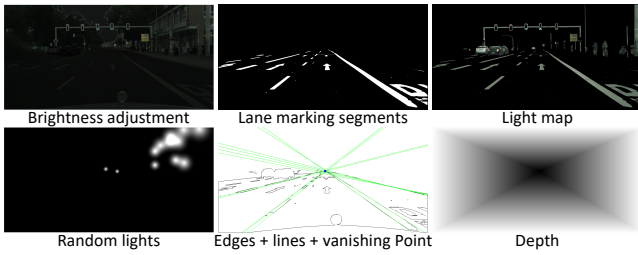
## 2. Our Approach

Our framework uses semantic annotation to extract scene elements from the input image. Hence, we perform the per-element translation and feature extractions during the translation. The overview of our framework is shown in Figure 1.

### 2.1. Semantic Adjustment of Brightness

A night-time image is generated by first adjusting the brightness of a day-time image, as the night scene lacks the sunlight. To only adjust the brightness of the image, we convert the image to HSV color space. But the brightness differs for each object in the scene, due to the variances in exposure. Hence, we perform per-element adjustments. The brightness of each element in the image is adjusted individually, based on the differences in exposure between the two domains. Thereby, a more adaptable generation of night-time images is feasible for our framework.

<sup>†</sup> Corresponding author



**Figure 2:** Example features used during the translation.

## 2.2. Depth-Aware Light Map Generation

To enhance the local contrast for lightings in the image, we generate a light map as follows. First, scene elements with specific semantic labels (i.e., traffic signs) are selected and automatically added to the light map, as they are well visible during the night. Likewise, lane markings are segmented [BSWG14]. For light sources, we first only consider traffic lights and vehicles, as the shape and location are easily recognizable (i.e., circular traffic lights and rectangular taillights). To further accurately estimate light sources from day-time images is yet a challenge even for the GAN-based methods, further perception of the scene with additional semantic segmentation is required for our framework. Hence, we splat light sprites randomly over selected scene elements where actual light sources are unpredictable (i.e., buildings). Prior to the final step, depth is also estimated. Thus, the light map can be applied to the night-time image with better awareness in depth and space, as the final result. For depth estimation, the following steps are performed.

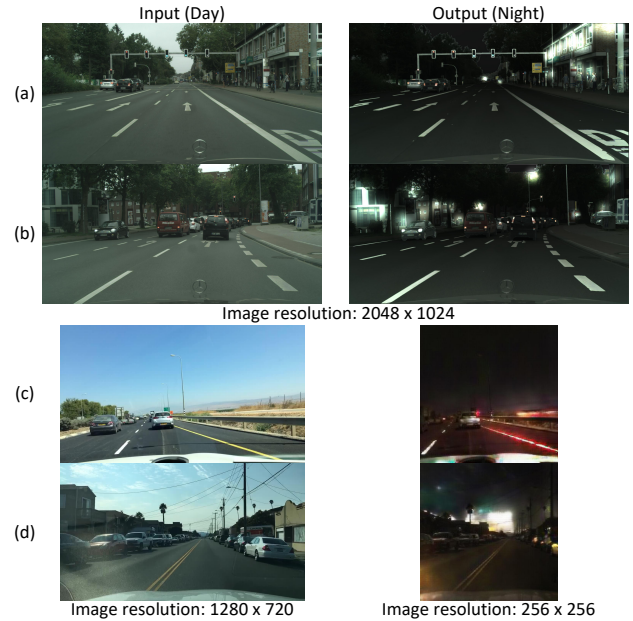
**Depth Estimation** Coarse depth is estimated based on the vanishing point. To obtain a vanishing point, edges are detected from selected semantic labels (i.e., road) and lane markings. Lines are detected from the edges and the largest point cluster [EKS\*96] is found from all intersection points of lines. The average point of the cluster is set to a vanishing point, which is set to the furthest depth.

## 3. Result

We tested our framework with the images from the Cityscapes [COR\*16] dataset. The results of our framework are shown in Figures 3(a) and 3(b), compared with the GAN-based method [APB\*19], as shown in Figures 3(c) and 3(d). By virtue of semantic annotation, our framework perceives the scene and extracts the features during the translation, and examples can be seen in Figure 2. Even with the addition of random lights, we achieved plausible results of high-resolution day-to-night domain adaptation through the per-element translation. The GAN-based method, in contrast, produces artifacts as lights are added into unpredictable locations (i.e. lane markings or sky). Also, the scene structures are difficult to observe due to the low resolution.

## 4. Limitations and Future Work

Our work has plenty of room for improvement. As our framework operates semi-automatically, it requires semantic segmentation of day-time images and additional adjustments are necessary for each



**Figure 3:** Results of our framework (a and b) compared to results of GAN-based method [APB\*19] (c and d).

translation. Also, light sources are randomly splatted, which is far from accurate light estimations. Our future work includes further investigation on fully automating the framework while estimating lights in the night scene with realistic optical effects applied.

## Acknowledgement

This work was supported in part by the Samsung Research Funding & Incubation Center of Samsung Electronics (SRFC-IT1901-01) and Mid-career R&D program through the NRF grants (2019R1A2C2002449) funded by the Korea Government. Correspondences concerning this article can be addressed to Sungkil Lee.

## References

- [APB\*19] ARRUDA V. F., PAIXÃO T. M., BERRIEL R. F., DE SOUZA A. F., BADUE C., SEBE N., OLIVEIRA-SANTOS T.: Cross-domain car detection using unsupervised image-to-image translation: From day to night. In *Proc. Int. J. Conf. Neural Networks* (2019), pp. 1–8. 2
- [BSWG14] BATISTA M. P., SHINZATO P. Y., WOLF D. F., GOMES D.: Lane detection and estimation using perspective image. In *Proc. Joint Conf. Robotics: SBR-LARS Robotics Symp. and Robocontrol* (2014), pp. 25–30. 2
- [COR\*16] CORDTS M., OMNAN M., RAMOS S., REHFELD T., ENZWEILER M., BENENSON R., FRANKE U., ROTH S., SCHIELE B.: The cityscapes dataset for semantic urban scene understanding. In *Proc. CVPR* (2016). 2
- [EKS\*96] ESTER M., KRIEGER H.-P., SANDER J., XU X., ET AL.: A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (1996), vol. 96, pp. 226–231. 2
- [IZZE17] ISOLA P., ZHU J.-Y., ZHOU T., EFROS A. A.: Image-to-image translation with conditional adversarial networks. In *Proc. of the IEEE Conf. on CVPR* (2017), pp. 1125–1134. 1