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

Seung Youp Baek Sungkil Lee

Sungkyunkwan University

Motivation

Day-to-night image translation?

A task that the goal is to translate the domain of an image from day to night. Recent studies approach with learning-based methods, but the results are yet unpredictable and difficult to be tuned for the desired output. Hence, we can approach without the learning process, and use the semantic annotation for a more adaptable and predictable solution.

Challenges

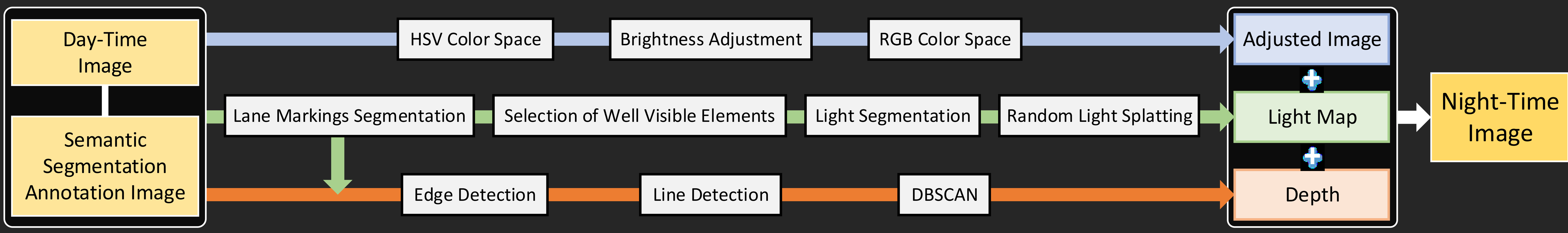
Day-to-night domain adaptation is considered as one of the most difficult domain to translate.

- Estimating lights in the night-time image based on the day-time image
- Acquiring datasets for training
- Tuning the results

Previous Approaches

Generative Adversarial Networks (GANs) based methods for the day-to-night image translation.

- Pix2pix [Isola et al. 2017]
  - Requires paired dataset of day and night
- [Arruda et al. 2019]
  - Observable artifacts in light estimations



Our Approach

We design a semi-automatic framework for the day-to-night image translation of road scenes. Unlike previous studies, the learning process is avoided for translation. Instead, we use the semantic segmentation annotation. This allows us to perform per-element translations and feature extractions on selected scene elements. Hence, unintentional translations are likely to be avoided and adjustments can be applied for each element in the scene.

Each step in our framework is described as follows.

1. Semantic Adjustment of Brightness

Adjust the brightness.

- Convert the image to HSV color space.
- Perform the per-element brightness adjustment to generate a night-time image.
- Convert the image back to RGB color space.

2. Depth-Aware Light Map Generation

Generate the light map.

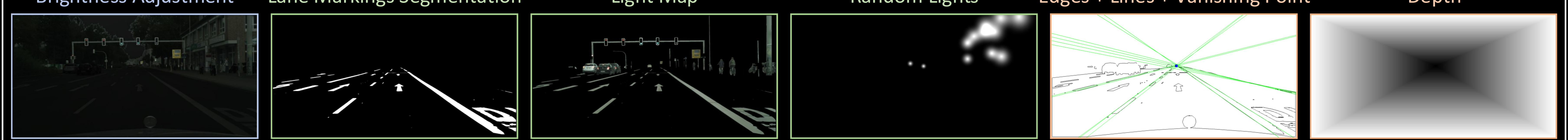
- Select well visible scene elements (i.e., traffic signs) and add to the light map.
- Segment the lane markings from the image and add to the light map.
- Manually segment the light sources from the traffic lights and vehicles, as the shape is easily recognizable (i.e., circular traffic lights).
- Splat random light sources over selected scene elements (i.e., buildings), where actual light sources are difficult to be estimated.

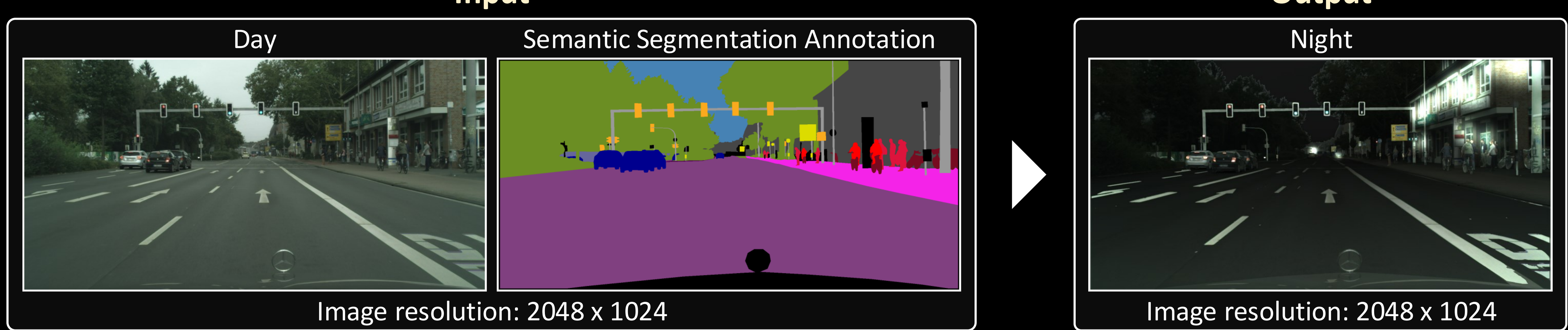
3. Depth Estimation

Estimate the coarse depth.

- Detect edges from selected scene elements (i.e., road) and lane markings as it well follows the perspective of the image.
- Detect lines from edges, list all intersection points between the lines.
- Find the largest point cluster, set the average point as the vanishing point.
- Coarse depth is generated by setting the vanishing point as the furthest depth.

The **final result** of our framework is generated by applying the **light map** over the **adjusted image** with the **depth** associated.





Experimental Results and Comparisons

Experimental results of our framework demonstrate high-resolution day-to-night image translation of road scenes. Even with the random lights splatted, our results are more plausible as we achieved the per-element translation with the aid of semantic annotation. In contrast, the learning-based methods have visible artifacts of estimated lights, and the scene structures are difficult to be observed, due to the low resolution.

Limitations

- Requires semantic segmentation annotation
- Manually segments light sources
- Randomly splats light sources
- Lacks optical effects