

Text-guided Transient Attribute Transfer

Keywords: transient attribute transfer, latent diffusion models

The appearance of a scene can change dramatically over a day or across seasons. While main elements, such as buildings, lakes, or forests generally remain the same, we often notice illumination changes that alter the appearance of the scene's contents. Additionally, we might observe content being added or removed between seasons or throughout the day. For example, a scene may be covered with snow in a transient attribute change from "summer" to "winter". These changes are difficult to estimate implicitly due to the entanglement of content, illumination, and reflectance. In this work, we explore conditional generative models for transferring transient attributes based on a target attribute. Specifically, we use variants of pre-trained latent diffusion models conditioned with text prompts indicating the target attribute.

We first fine-tune a stable diffusion model using the v1.5 checkpoint weights (Rombach et al., 2022) and ControlNet (Zhang et al., 2023) on the Transient Attribute Dataset (Laffont al., 2014) of 8,571 images. However, we observe that these highly complex models can easily overfit to a small dataset even in a fine-tuning setting. This challenge leads us to explore a zero-shot setup that does not require an additional dataset, using these pre-trained foundation models as a prior to guide the desired transfer. We evaluate the mentioned techniques qualitatively on the test images from the Transient Attribute Dataset (Laffont al., 2014). The zero-shot approach effectively transfers the attributes while preserving high-level content.

