SatSynth: Augmenting Image-Mask Pairs through Diffusion Models for Aerial Semantic Segmentation

Problem they are trying to solve / Purpose of method

For semantic segmentation tasks, there is not that much annotated data. That is an issue as annoted data is pivotal for semantic segmentation models. Improving data labeling can help tackling various humanitarian challenges.

SatSynth proposes a method for generating synthetic satellite images with segmentation masks.

How does it differ from other methods?

Traditional approaches use General Adverserial Networks (GANs) for generating synthetic images and masks, or just use conventional image augmentations like flipping, cropping, and rotations.

SatSynth utilizes a diffusion process.

How the method works

The method can essentially be boiled down to using a diffusion model to generate/sythesize new images and label masks.

The core idea is to utilize a Denoising Diffusion Probabilistic Model (DDPM) to:

1. Learn the joint data distribution $p(x, y)$ of images x and labels y via a diffusion model G.
2. Then we utilize G to generate new training samples x' and y' as data augmentation.
3. Then we evaulate the method as usual.