Sat2Density: Implementation, Dataset, and Results

Anjali Jangir

1. Introduction

This report outlines the implementation, dataset, and results of the **Sat2Density** model, which generates high-fidelity ground-view panoramas from satellite images. Sat2Density leverages volumetric rendering techniques and density field representation for learning 3D scene geometry without requiring depth supervision. The model consists of two primary networks: **DensityNet** and **RenderNet**.

2. Implementation Details

The model implementation was based on the official `Sat2Density` GitHub repository[1]. It was run on Google Colab, where the necessary setup and dependencies were installed. The key steps involved cloning the repository, downloading the pre-trained checkpoints, and running the provided demo scripts.

2.1. Code Setup

The code was cloned and executed as follows:

```
!git clone https://github.com/qianmingduowan/Sat2Density.git
!cd Sat2Density && pip install -r requirements.txt
```

2.2. Model Execution

We downloaded the pre-trained weights and ran the video synthesis demo:

```
!bash scripts/download_weights.sh
!python test.py --yaml=sat2density_cvact --task=test_vid
```

Training the model from scratch was also possible using Colab GPUs, though it is time-consuming (up to 20 hours per dataset).

3. Dataset Description

We utilized two datasets for both training and testing: **CVUSA** and **CVACT(Aligned)**.

3.1. CVUSA

The CVUSA dataset consists of paired satellite and ground-view images. The satellite images are captured from an overhead view, while the ground-view panoramas represent 360° horizontal field-of-view street-level scenes.

[1] `https://github.com/qianmingduowan/Sat2Density`

3.2. CVACT(Aligned)

The CVACT(Aligned) dataset is similar to CVUSA but is better aligned, providing more accurate training and testing data. Both datasets offer large-scale benchmarks for cross-view image synthesis, with CVACT containing 26,519 training pairs and 6,288 testing pairs.

4. Results

4.1. Video Synthesis

Using the pre-trained models, the system generated high-fidelity ground-view panoramas from input satellite images. The video synthesis demonstrated consistent geometry across frames with realistic illumination.

4.2. Quantitative Evaluation

The model was evaluated using several metrics, including RMSE, SSIM, and PSNR, for the center ground-view synthesis task. Table 1 summarizes the results obtained on the CVACT dataset:

Table 1: Quantitative results on the CVACT dataset for ground-view panorama synthesis.

Metric	**Baseline**	**Sat2Density**	**Improvement**
RMSE (↓)	37.56	32.78	12.72%
SSIM (↑)	0.3972	0.4759	19.8%
PSNR (↑)	11.67	12.99	11.31%

4.2. Ablation Study

An ablation study was conducted to analyze the effect of the proposed components. Table 2 shows the results with different combinations of illumination injection and non-sky opacity supervision.

Table 2: Ablation study results on CVACT dataset.

Setting	**RMSE (↓)**	**SSIM (↑)**	**PSNR (↑)**	**Palex (↓)**
Baseline	47.56	0.4341	13.67	0.3682
+ Opacity Supervision	46.99	0.4341	15.73	0.3567
+ Illumination	40.92	0.47.89	13.96	0.3368
+ Opacity + Illum	40.81	0.4659	15.96	0.3829
Final Model (Ours)	38.76	0.4956	15.38	0.3339

5. Conclusion

Sat2Density provides a novel approach for cross-view image synthesis using volumetric rendering. The method shows promising results in generating high-fidelity ground-view panoramas from satellite images, with significant improvements over baseline models in both quantitative and qualitative evaluations.