GeoSynth: Contextually-Aware High-Resolution Satellite Image Synthesis

Supplementary Material

A. Qualitative Results on UCMerced

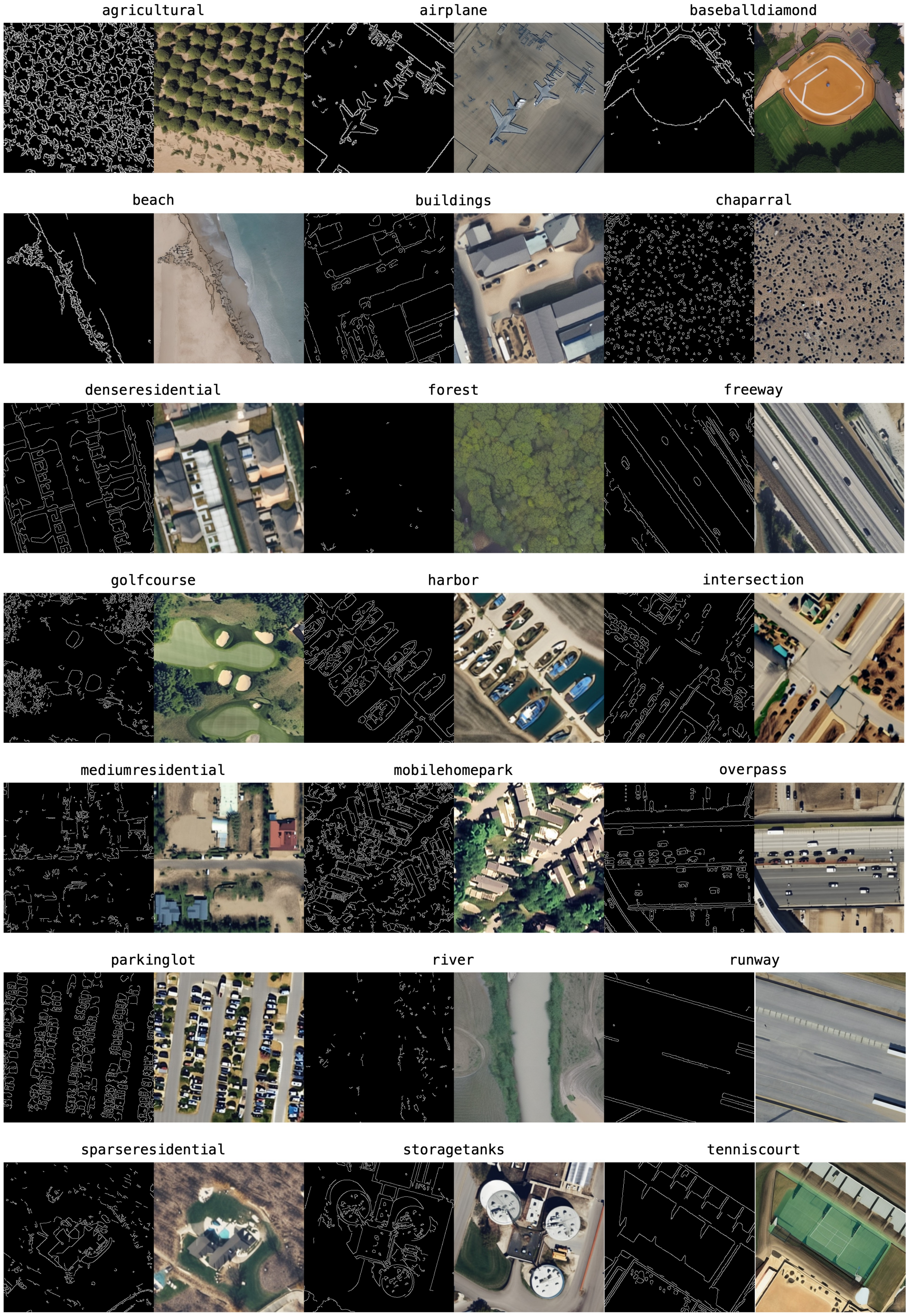

Figure 1. Zero-shot synthesis using Canny edges. Examples of synthesized satellite images using Canny edges generated on UCMerced dataset.

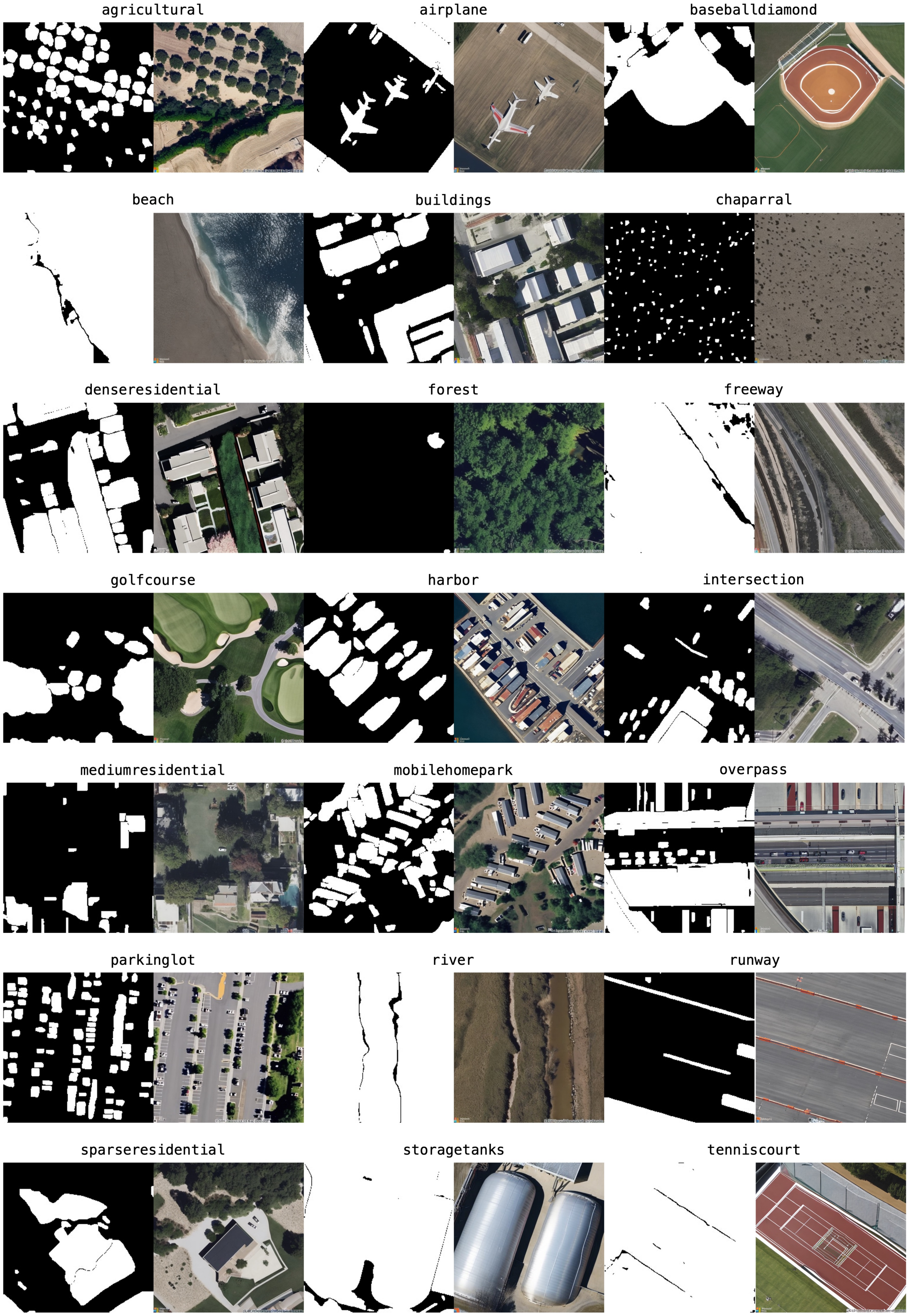

Figure 2. Zero-shot synthesis using SAM mask. Examples of synthesized satellite images using SAM mask generated on UCMerced dataset.

B. Dataset Sampling

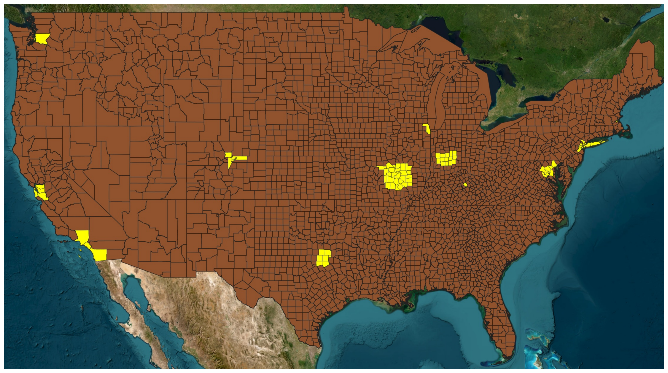

Figure 3. Dataset sampling. The geographic locations across the USA used for sampling the dataset.

C. Limitations

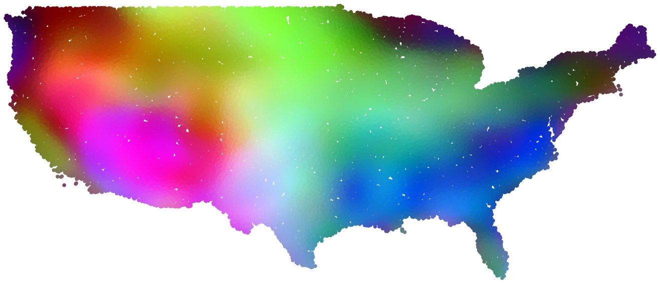

Figure 4. Features Learned by SatCLIP. We show the visualization of location embeddings learned by SatCLIP. For visualization, the embeddings are projected to a 3-dimensional space using Independent Component Analysis.

GeoSynth is a general framework for synthesizing satellite images that combines various state-of-the-art components. As a result, its synthesis performance highly depends on the performance of each of the components. In particular, the geo-awareness of our model is restricted by the ability of SatCLIP to represent the world's geography effectively. In Figure 4, we show an Independent Component Analysis (ICA) plot of SatCLIP embeddings over the USA. The figure depicts that SatCLIP cannot capture high-frequency information about the world's geography at a fine scale. Given the flexibility of our framework, SatCLIP can easily be replaced with another location encoder in the future. Currently, GeoSynth only supports synthesizing RGB-based satellite images. It is possible to extend the architecture of GeoSynth for synthesizing images coming from different modalities such as depth, radar, etc.

D. Acknowledgements

We thank Dr. Joseph O'Sullivan, Wiete Fehner, and others for useful discussions as part of ESE 5934. The code for this project is highly inspired from Zhang *et al.* [1] GitHub repository: `https://github.com/lllyasviel/ControlNet`.

References

[1] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 3