# Generative Gaussian Splatting for Unbounded 3D City Generation
# Supplementary Material

Haozhe Xie, Zhaoxi Chen, Fangzhou Hong, Ziwei Liu ✉
S-Lab, Nanyang Technological University
{haozhe.xie, zhaoxi001, fangzhou.hong, ziwei.liu}@ntu.edu.sg
https://haozhexie.com/project/gaussian-city

In this supplementary material, we supplement the following materials to support the findings and conclusions drawn in the main body of this paper.

# A. Additional Ablation Study Results

## A.1. Qualitative Results for Ablation Studies

**Effectiveness of BEV-Point Decoder.** Figure I gives a qualitative comparison as a supplement to Table 3, demonstrating the effectiveness of Point Serializer and Point Transformer in BEV-Point Decoder. Removing either of them significantly degrades the quality of the generated images.

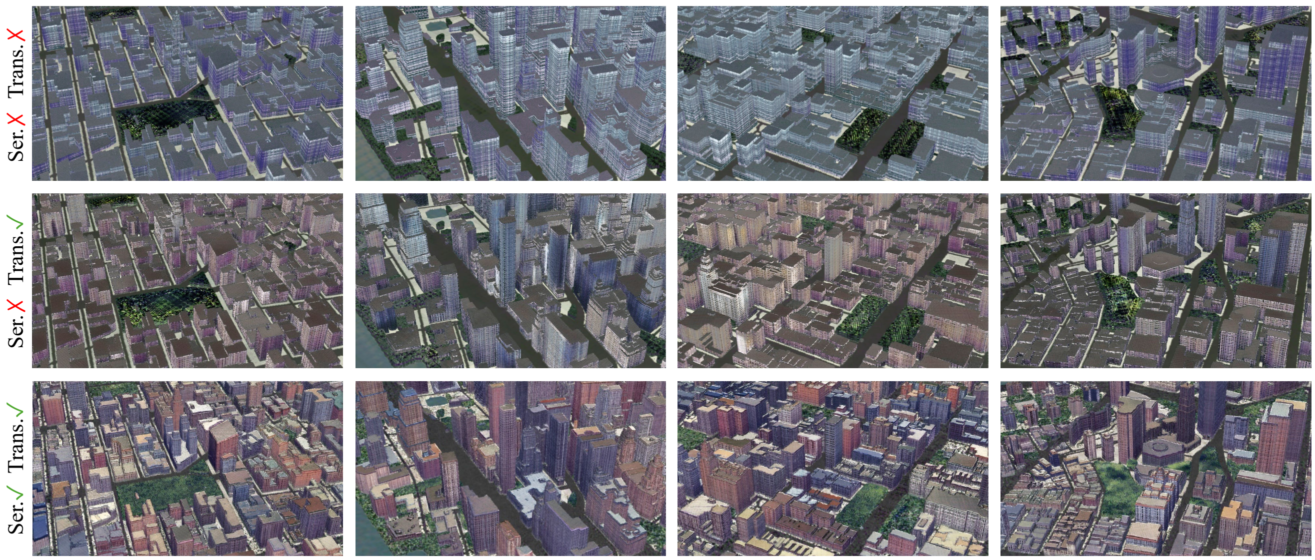


Figure I. **Qualitative comparison of different BEV-Point Decoder variants.** Note that "Ser." and "Trans." represent "Point Serializer" and "Point Transformer", respectively.

**Effectiveness of Different Serialization Methods.** Figure II provides a qualitative comparison as a supplement to Table 4. As shown in Figure II, using either Equation 13, the Hilbert curve, or both for serialization is unlikely to impact the quality of the generated results. Moreover, Equation 13 exhibits significantly lower computational complexity compared to the Hilbert curve. Therefore, Equation 13 is employed in the Point Serializer due to its lower computational demands.

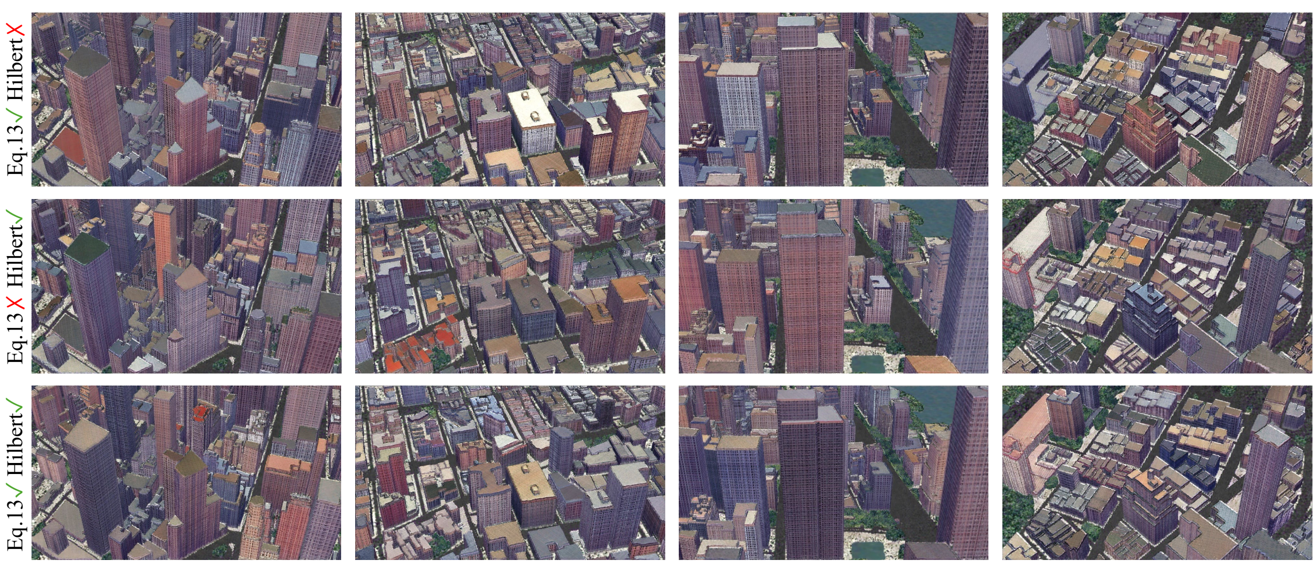


Figure II. **Qualitative comparison of different serialization methods.** Note that "Eq. 13" denote the serialization used in Point Serializer.

### A.2. Discussion on Generated Gaussian Attributes

In 3D-GS [2], each 3D Gaussian possesses multiple attributes, including XYZ coordinates, spherical harmonics (SHs), opacity, rotation, and scale. These attributes are optimized using supervision from multi-view images to represent the scene. In reconstruction, adding extra attributes like XYZ offsets and opacity enhances representation capability. However, a scene can have multiple valid reconstruction results because different combinations of 3D Gaussian attributes can yield the same rendering result. For instance, changing the color of one 3D Gaussian can be equivalent to overlaying multiple Gaussians with varying opacities or adjusting the scale of nearby Gaussians with similar colors. This ambiguity causes instability when optimizing all these attributes simultaneously in city generation.

Due to the carefully designed BEV-Point initialization, all points are evenly distributed on the surface of objects. Therefore, Gaussian attributes other than RGB can be left unestimated and set to their default values. Table I and Figure III provide a quantitative and qualitative comparison of various attributes generated for 3D Gaussians, demonstrating that optimizing 3D Gaussian attributes beyond RGB not only complicates network convergence but also significantly impacts the quality of the generated results.

Table I. **Quantitative comparison of different attributes generated for 3D Gaussians.** The best results are highlighted in bold. Note that "-" in CE indicates that COLMAP cannot estimate camera poses from the generated images.

| Generated Attributes | | | | GoogleEarth | | | | KITTI-360 | |
|---|---|---|---|---|---|---|---|---|---|
| RGB | ΔXYZ | Opacity | Scale | FID ↓ | KID ↓ | DE ↓ | CE ↓ | FID ↓ | KID ↓ |
| ✓ | ✗ | ✗ | ✗ | **86.94** | **0.090** | **0.136** | **0.057** | **29.5** | **0.017** |
| ✓ | ✓ | ✗ | ✗ | 371.15 | 0.468 | 0.367 | - | 281.5 | 0.342 |
| ✓ | ✓ | ✓ | ✗ | 384.81 | 0.485 | 0.401 | - | 256.9 | 0.297 |
| ✓ | ✓ | ✓ | ✓ | 535.49 | 0.709 | 0.470 | - | 486.3 | 0.605 |

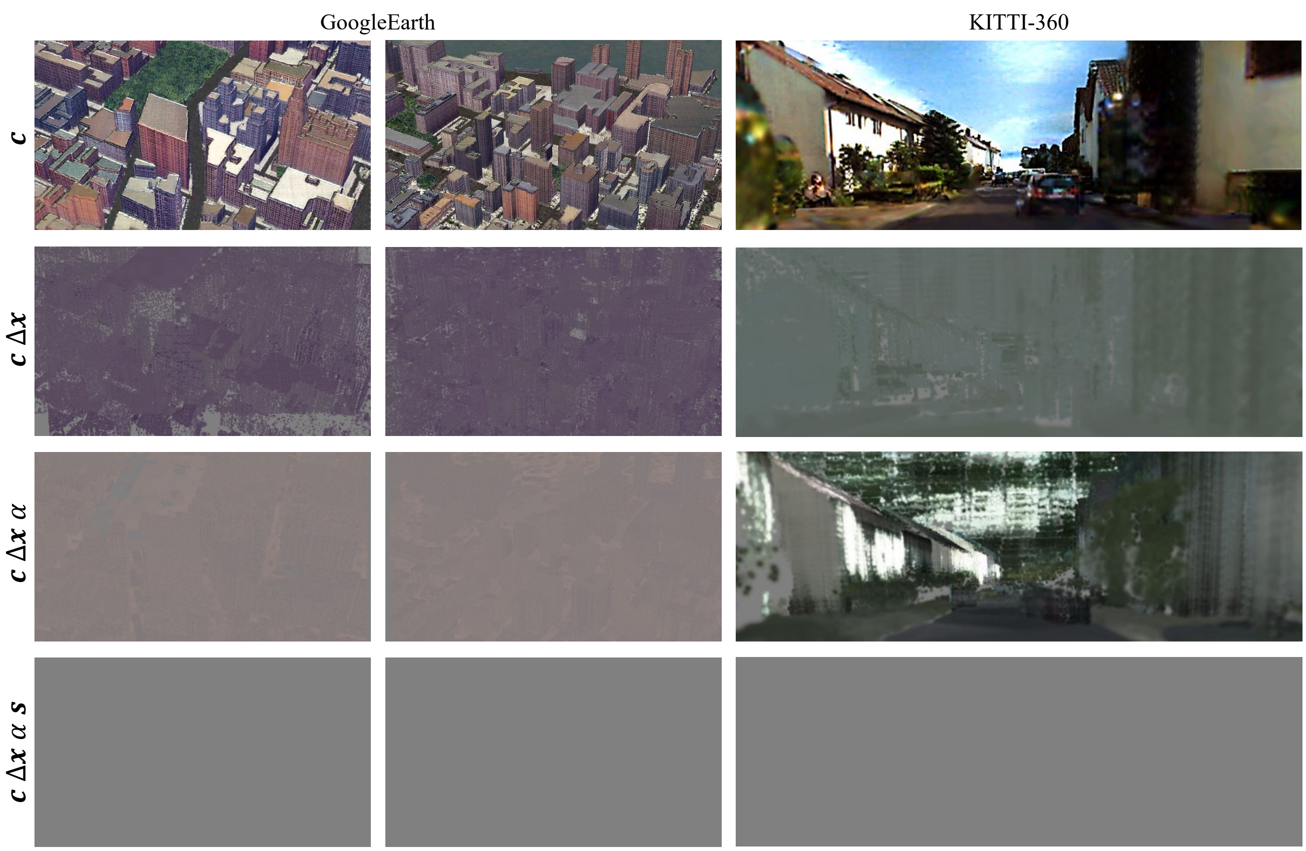


Figure III. **Qualitative comparison of different attributes generated for 3D Gaussians.** Note that $\mathbf{c}$, $\Delta\mathbf{x}$, $\alpha$, and $\mathbf{s}$ denote the RGB color, XYZ offsets, opacity, and scale in the generated 3D Gaussian attributes, respectively.

## B. Additional Experimental Results

### B.1. Building Interpolation

As shown in Figure IV, GaussianCity showcases the capability to interpolate building styles controlled by the variable $z$. In the first row, local editing is applied by altering only the style code $z$ of the buildings within the bounding boxes, leaving the styles of the other buildings unchanged. In the second row, the style codes of all the buildings are interpolated from left to right, resulting in a gradual style transition across the entire row.

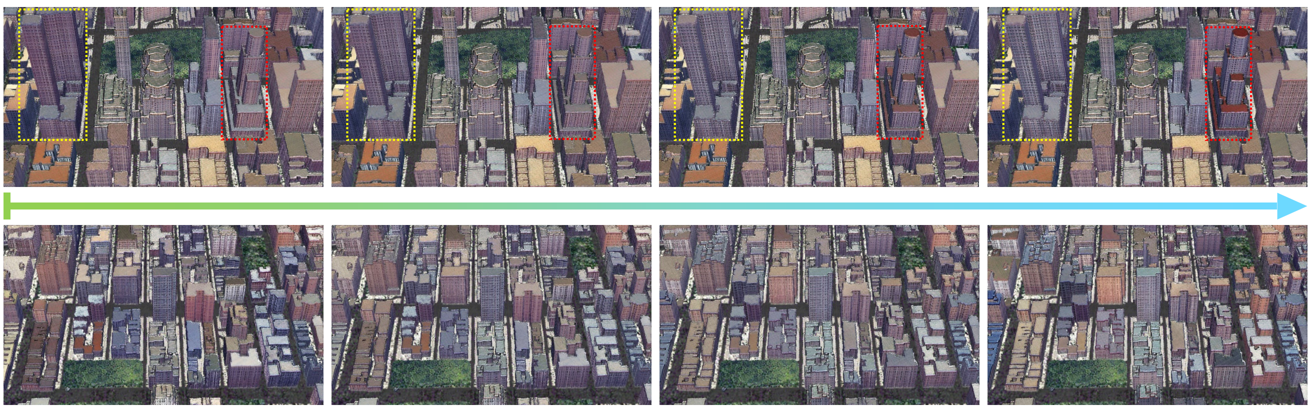

Figure IV. **Linear interpolation along the building style.** The style of each building gradually changes from left to right. In the first row, only the styles of the buildings enclosed by the bounding boxes are modified, while in the second row, the styles of all the buildings are changed.

### B.2. Relighting

From the explicit representation of 3D Gaussians, relighting is much simpler than NeRF-based methods. Using Luma AI's 3D Gaussians Plugin [1], the generated city can be imported into Unreal Engine, enabling highly realistic lighting effects, as illustrated in Figure V. However, the plugin is still under development, so shadow effects are not well generated. We believe this issue will be resolved soon.

Figure V. **Relighting effects in Unreal Engine 5.** From left to right, the relighting effect is shown with increasing light intensity.

[1] https://www.unrealengine.com/marketplace/product/66dd775fa3104ecfb3ae800b8963c8b9

### B.3. Additional Qualitative Comparison

In Figures VI, VII and VIII, we provide more visual comparisons with state-of-the-art methods on GoogleEarth and KITTI-360, respectively.

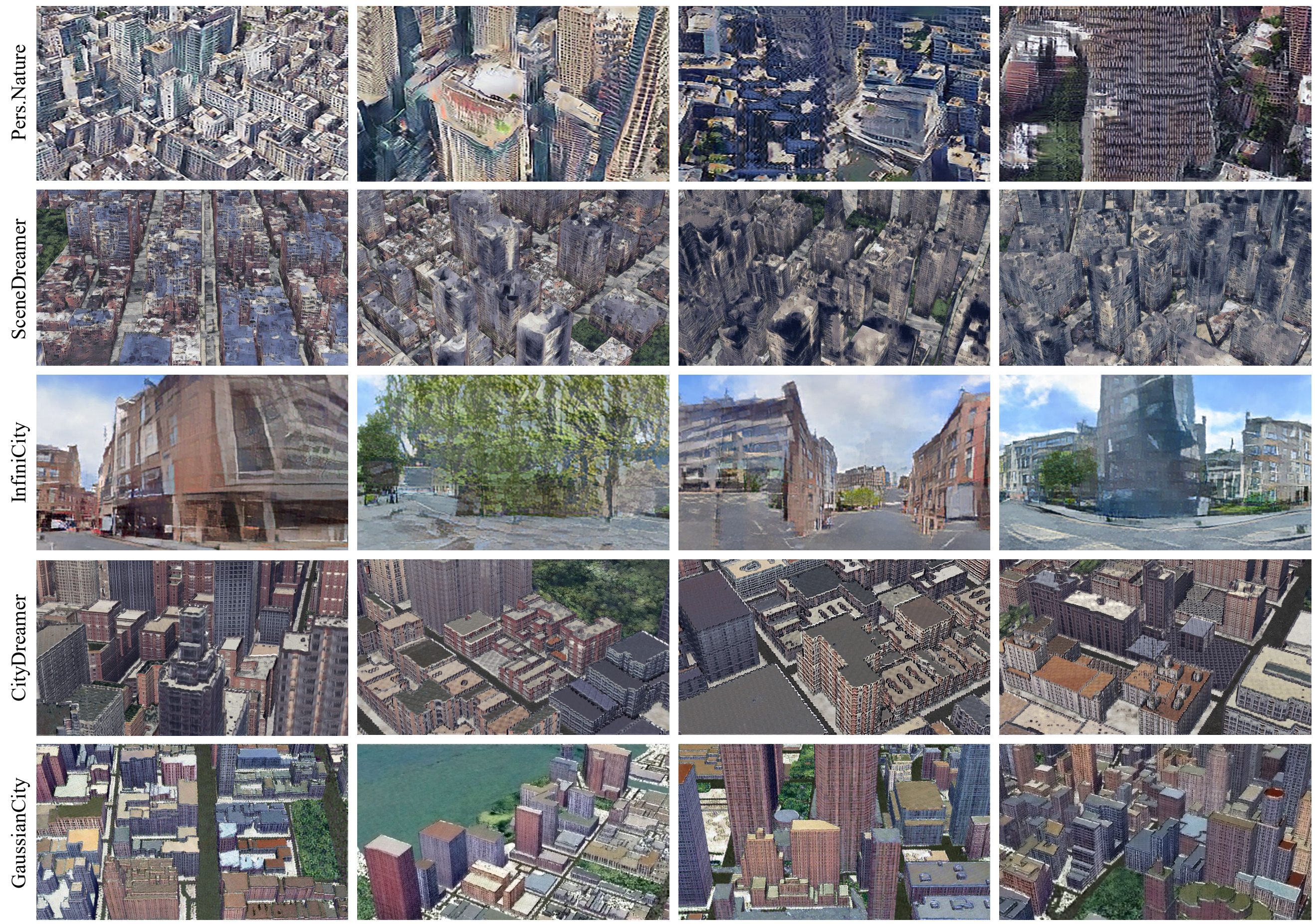


Figure VI. **Qualitative comparison on GoogleEarth.** Note that "Pers.Nature" is short for "PersistentNature" [1]. The visual results of InfiniCity [3] are provided by the authors and zoomed for optimal viewing.

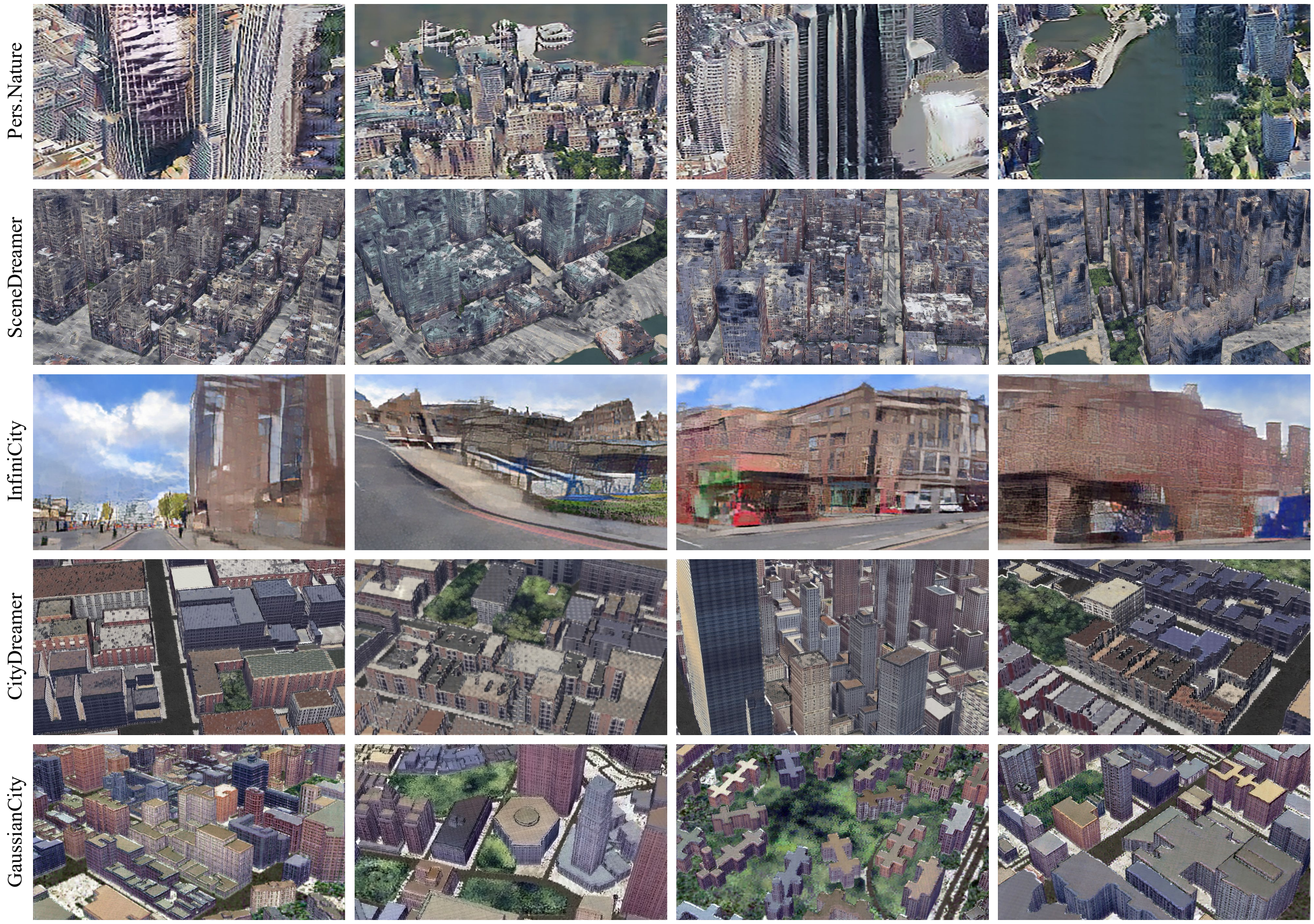


Figure VII. **Qualitative comparison on GoogleEarth.** Note that "Pers.Nature" is short for "PersistentNature" [1]. The visual results of InfiniCity [3] are provided by the authors and zoomed for optimal viewing.


Figure VIII. **Qualitative comparison on KITTI-360.** The visual results of UrbanGIRAFFE [4] are provided by the authors since the training code and pretrained model are unavailable.

## References

[1] Lucy Chai, Richard Tucker, Zhengqi Li, Phillip Isola, and Noah Snavely. Persistent Nature: A generative model of unbounded 3D worlds. In *CVPR*, 2023. 5, 6

[2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM TOG*, 42(4):139:1–139:14, 2023. 3

[3] Chieh Hubert Lin, Hsin-Ying Lee, Willi Menapace, Menglei Chai, Aliaksandr Siarohin, Ming-Hsuan Yang, and Sergey Tulyakov. InfiniCity: Infinite-scale city synthesis. In *ICCV*, 2023. 5, 6

[4] Yuanbo Yang, Yifei Yang, Hanlei Guo, Rong Xiong, Yue Wang, and Yiyi Liao. UrbanGIRAFFE: Representing urban scenes as compositional generative neural feature fields. In *ICCV*, 2023. 7