CityDreamer

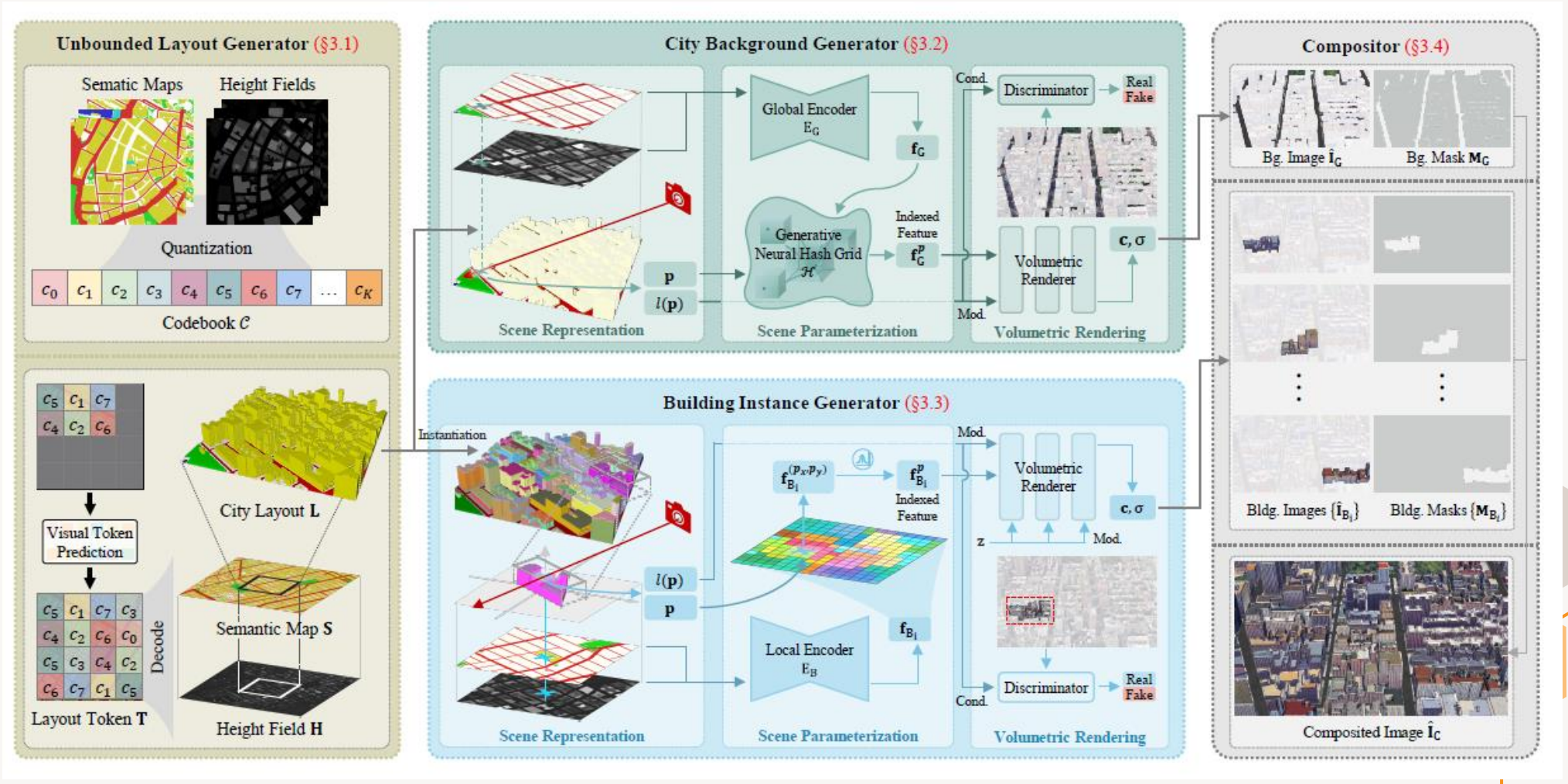
Fundamentals and Trends in Vision and Image Processing

Professor: Luiz Velho (IMPA)
Aluno: Daniel Morais (IME)

Definition

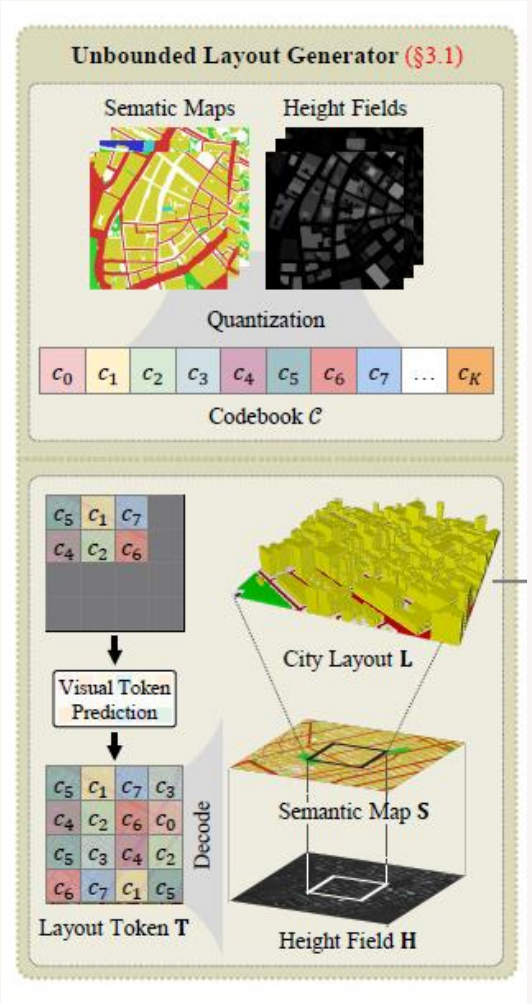
- CityDreamer is a compositional generative model specifically designed for creating unbounded 3D cities.
- It separates the generation of building instances from other background objects, such as roads, green lands, and water areas, into distinct modules.
- This unique strategy of isolating the creation of building instances from other background elements often present in cities sets it apart.
- CityDream uses 2 datasets:
  - Google Earth
  - OpenStreetMap (OSM)

Overview


Unbounded Layout Generator (§3.1)
Sematic Maps
Height Fields
Quantization
Codebook $\mathcal{C}$
Visual Token Prediction
Layout Token T
Decode
City Layout L
Semantic Map S
Height Field H
City Background Generator (§3.2)
Global Encoder E_G
Generative Neural Hash Grid $\mathcal{H}$
Indexed Feature
Scene Representation
Scene Parameterization
Volumetric Rendering
Discriminator
Real Fake
Cond.
Mod.
Building Instance Generator (§3.3)
Instantiation
Local Encoder E_B
Compositor (§3.4)
Bg. Image $\hat{I}_G$
Bg. Mask M_G
Bldg. Images $\{\hat{I}_{B_i}\}$
Bldg. Masks $\{M_{B_i}\}$
Composited Image $\hat{I}_C$

Generation Steps

- Initially, the **unbounded layout generator** creates an arbitrary large city layout L.



Volume Function:

$$L_{(i,j,k)} = \begin{cases} S_{(i,j)} & \text{if } k \leq H_{(i,j)} \\ 0 & \text{otherwise} \end{cases}$$

Smoothness Loss Function:

$$\ell_{\mathrm{VQ}} = \lambda_{\mathrm{R}} \|\hat{\mathbf{H}}_{\mathbf{p}} - \mathbf{H}_p\| + \lambda_{\mathrm{S}} \mathcal{S}(\hat{\mathbf{H}}_{\mathbf{p}}, \mathbf{H}_p) + \lambda_{\mathrm{E}} \mathcal{E}(\hat{\mathbf{S}}_{\mathbf{p}}, \mathbf{S}_p)$$

Generation Steps

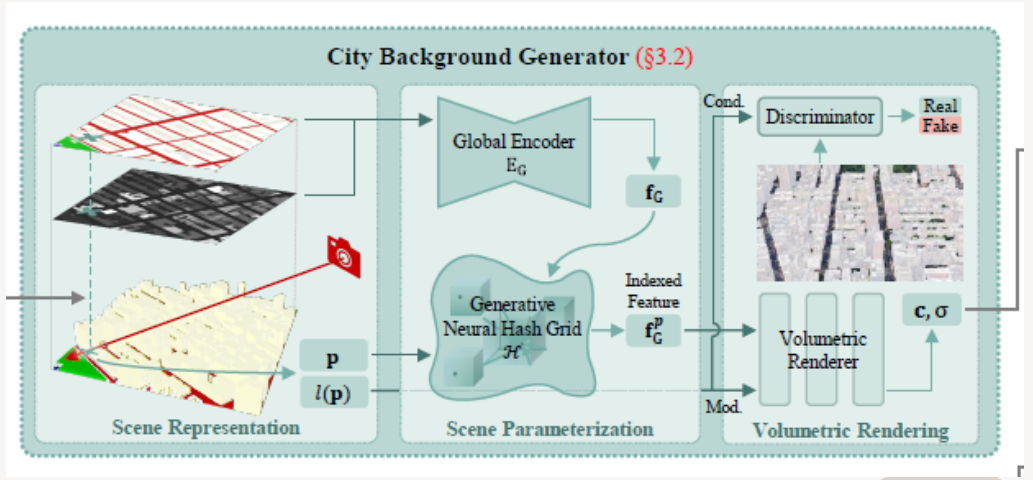
- Subsequently, the city **background generator** produces the background image ($\hat{I}_G$) along with its corresponding mask (M_G).

Compact city level feature:

$$\mathbf{f}_\mathrm{G} = E_\mathrm{G}(\mathbf{H}_\mathrm{G}^\mathrm{Local}, \mathbf{S}_\mathrm{G}^\mathrm{Local})$$

Indexed “f” at 3D position “p”:

$$\mathbf{f}_\mathrm{G}^\mathrm{p} = \mathcal{H}(\mathbf{p}, \mathbf{f}_\mathrm{G}) = \left(\bigoplus_{i=1}^{d_G} f_\mathrm{G}^i \pi^i \bigoplus_{j=1}^{3} p^j \pi^j \right) \mod T$$



Volumetric Rendering:

$$C(\mathbf{r}) = \int_0^\infty T(t)\mathbf{c}(\mathbf{f}_\mathrm{G}^{\mathbf{r}(t)}, l(\mathbf{r}(t)))\boldsymbol{\sigma}(\mathbf{f}_\mathrm{G}^{\mathbf{r}(t)})dt$$

Loss Function:

$$\ell_\mathrm{G} = \lambda_{\mathrm{L1}}\|\hat{\mathbf{I}}_\mathrm{G} - \mathbf{I}_\mathrm{G}\| + \lambda_\mathrm{P}\mathcal{P}(\hat{\mathbf{I}}_\mathrm{G}, \mathbf{I}_\mathrm{G}) + \lambda_\mathrm{G}\mathcal{G}(\hat{\mathbf{I}}_\mathrm{G}, \mathbf{S}_\mathrm{G})$$

Generation Steps

- Next, the **building instances generator** generates images for building instances and their respective masks, where n is the number of building instances.

Pixel-level feature:

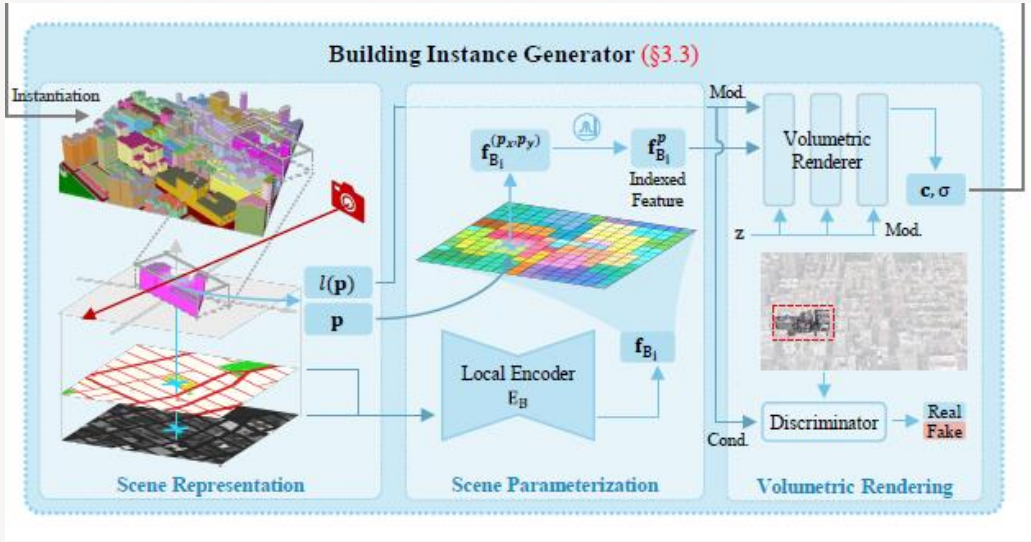
$$\mathbf{f}_{\mathrm{B_i}} = E_B(\mathbf{H}_{\mathrm{B_i}}^{\mathrm{Local}}, \mathbf{S}_{\mathrm{B_i}}^{\mathrm{Local}})$$

P indexed feature computed:

$$\mathbf{f}_{\mathrm{B_i}}^{\mathrm{P}} = \mathcal{O}(\mathrm{Concat}(\mathbf{f}_{\mathrm{B_i}}^{(p_x,p_y)}, p_z))$$

Volumetric Rendering:

$$C(\mathbf{r}) = \int_0^{\infty} T(t)\mathbf{c}(\mathbf{f}_{\mathrm{B_i}}^{\mathbf{r}(t)}, \mathbf{z}, l(\mathbf{r}(t)))\boldsymbol{\sigma}(\mathbf{f}_{\mathrm{B_i}}^{\mathbf{r}(t)})dt$$



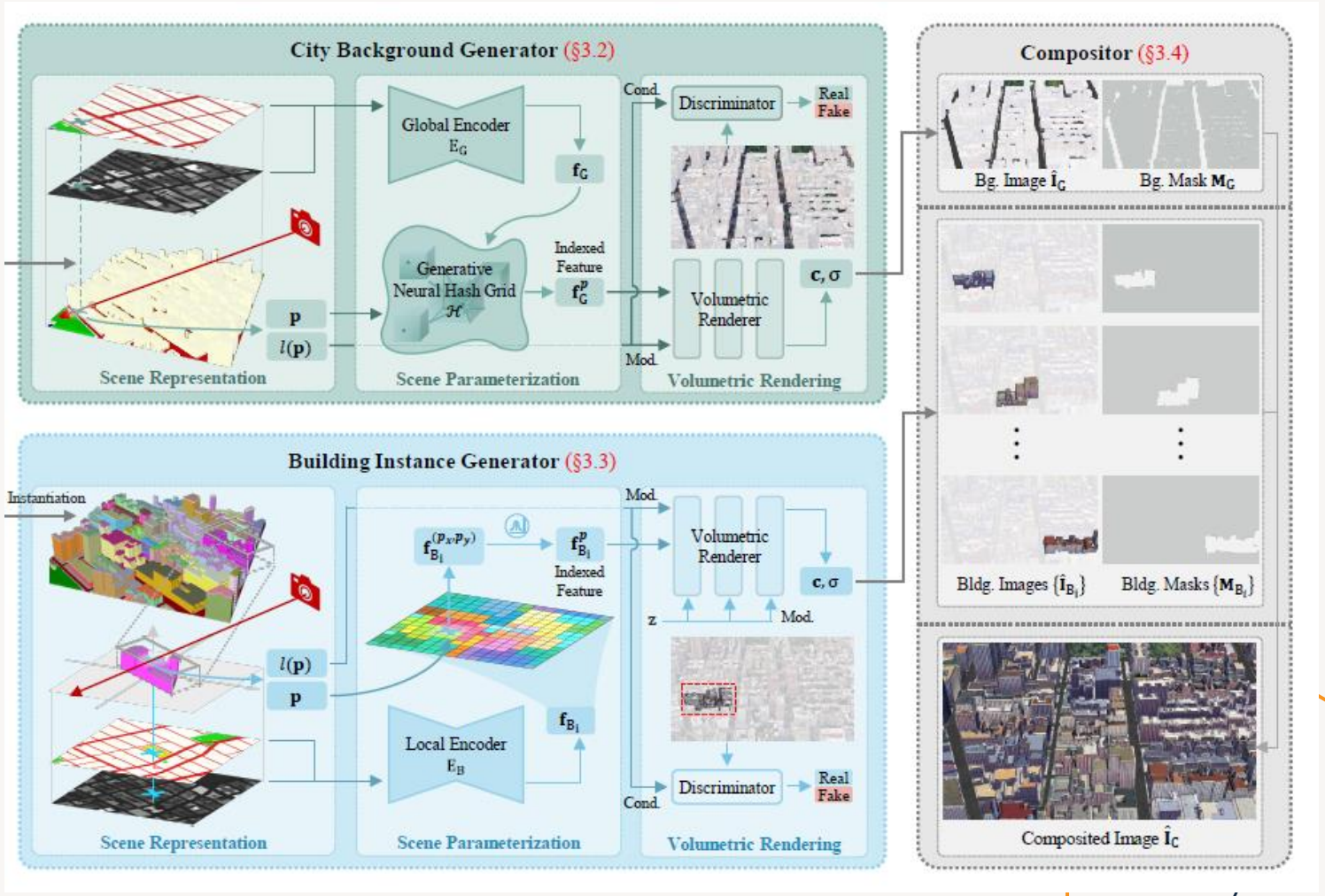
Gan loss function:

$$\ell_{\mathrm{B}} = \mathcal{G}(\hat{\mathbf{I}}_{\mathrm{B_i}}, \mathbf{S}_{\mathrm{B_i}})$$

Generation Steps

- Lastly, the **compositor** merges the rendered background and building instances into a single cohesive image IC.

$$\mathbf{I}_C = \hat{\mathbf{I}}_G \mathbf{M}_G + \sum_{i=1}^{n} \hat{\mathbf{I}}_{B_i} \mathbf{M}_{B_i}$$



Benchmarks

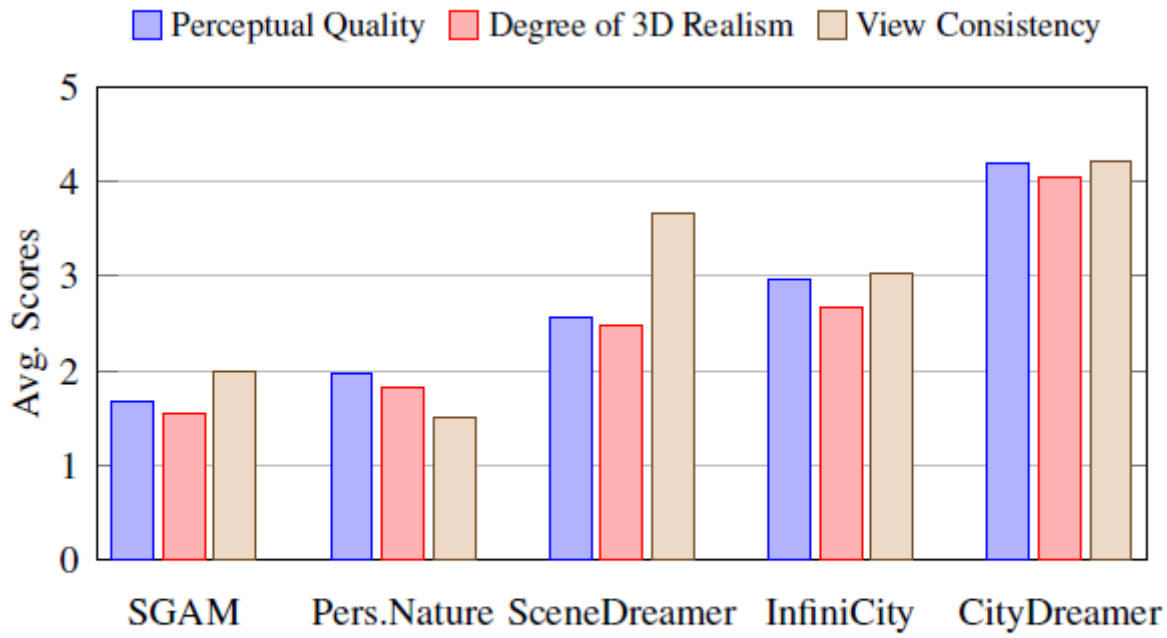


Figure 5. **User study on unbounded 3D city generation.** All scores are in the range of 5, with 5 indicating the best.

Benchmarks

Table 3. **Effectiveness of Ubounded Layout Generator.** The best values are highlighted in bold. The images are centrally cropped to a size of 4096×4096.

Methods	FID ↓	KID ↓
IPSM [10]	321.47	0.502
InfinityGAN [33]	183.14	0.288
Ours	**124.45**	**0.123**

Table 4. **Effectiveness of Building Instance Generator.** The best values are highlighted in bold. Note that "w/o BIG." indicates the removal of Building Instance Generator from CityDreamer. "w/o Ins." denotes the absence of building instance labels in the Building Instance Generator.

Methods	FID ↓	KID ↓	DE ↓	CE ↓
w/o BIG.	213.56	0.216	0.152	0.186
w/o Ins.	117.75	0.124	0.148	0.098
Ours	**97.38**	**0.096**	**0.147**	**0.060**

Table 5. **Effectiveness of different generative scene parameterization.** The best values are highlighted in bold. Note that "CBG." and "BIG." denote City Background Generator and Building Instance Generator, respectively. "Enc." and "P.E." represent "Encoder" and "Positional Encoding", respectively.

CBG.		BIG.		FID ↓	KID ↓	DE ↓	CE ↓
Enc.	P.E.	Enc.	P.E.				
Local	SinCos	Global	Hash	219.30	0.233	0.154	0.452
Local	SinCos	Local	SinCos	107.63	0.125	0.149	0.078
Global	Hash	Global	Hash	213.56	0.216	0.153	0.186
Global	Hash	Local	SinCos	**97.38**	**0.096**	**0.147**	**0.060**

CityDreamer

Compositional Generative Model of Unbounded 3D Cities

[illegible]

Obrigado!