MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency

Xingbo YAO, Xuanmin WANG, Hao WU, Chengliang PING, Doudou Zhang, Hui XIONG*

1.Introduction of our work

Task: Generate 3D urban scene on a given satellite image.

Why important: Games, urban simulation, and mapping services; transform real-world into digital twins.

2. How does previous work do?

(1) Geometry Prior-based Methods:

- InfiniCity
- CityDreamer
- GaussianCity

❌ *Lack style diversity and texture quality*

(2) Image Prior-based Methods:

- CAT3D
- DimensionX
- DreamScene

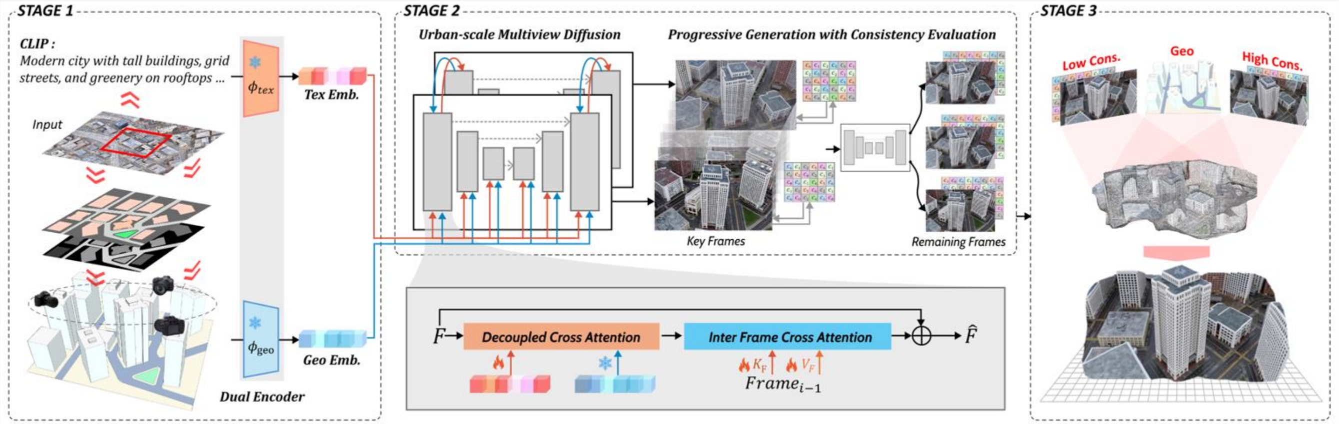
❌ *Lack geo. consistency in large scenes*

Motivation: How to achieve high texture quality while maintain geometric consistency?

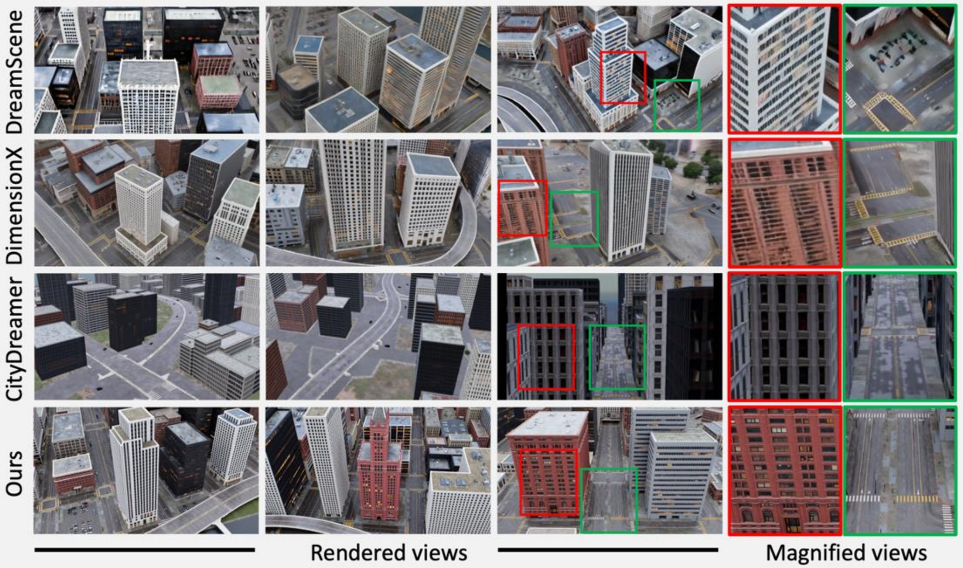
3. Our methods

Pipeline overview: Scene Initialization -> Multiview images Generation -> Robust Reconstruction

Insights: Multiview images gen. with *city-level consistency* ; Robust 3D reconstruction on *generated images*



4.Experimental Results



Dataset: CityVista, consisting of 500 high-quality city scenes.

Metrics: Texture quality & Geometric consistency.

Qualitative: Our method produces *higher-quality* 3D cities with *better consistency* compared with the baselines.

Quantitative: MagicCity outperforms the baselines *across all metrics.*

	CityDreamer	DimensionX	DreamScene	Ours
FID	155.390	126.890	104.627	**86.096**
KID	0.251	0.175	0.125	**0.087**
NIQE	8.632	6.595	6.018	**4.553**
BRISQUE	86.773	70.207	30.311	**28.018**
DE	0.157	-	0.223	**0.137**
CE	0.083	-	0.371	**0.072**

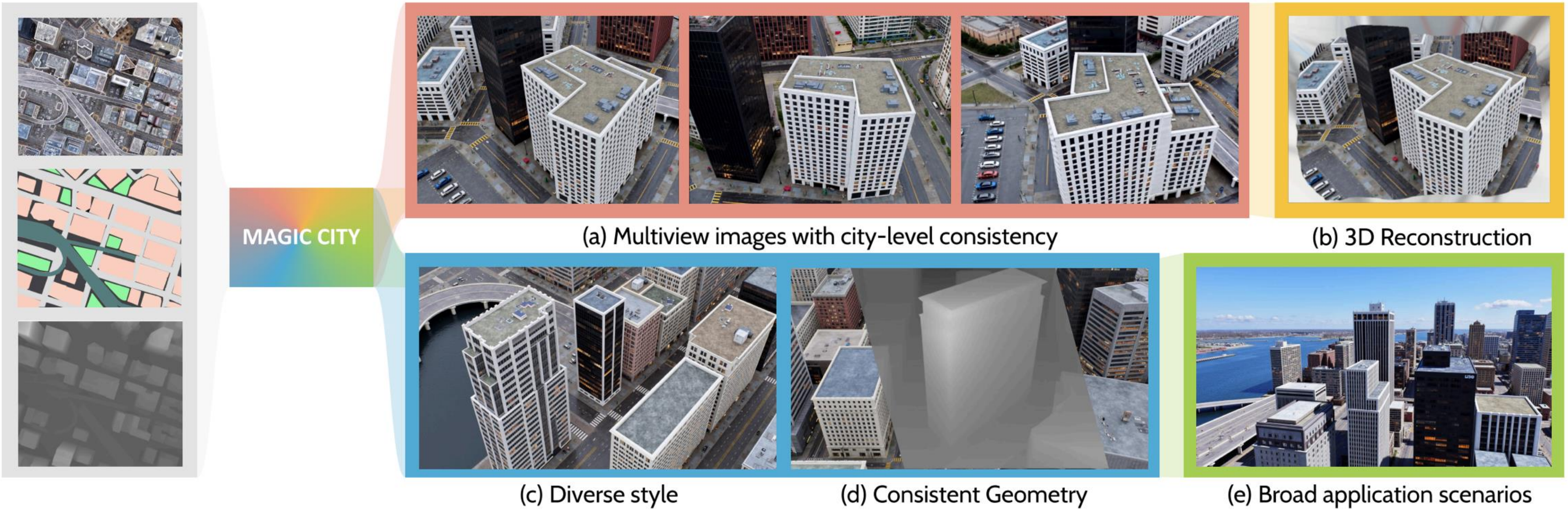
This work was supported in part by the National Key R&D Program of China (Grant No.2023YFF0725001), in part by the National Natural Science Foundation of China (Grant No.92370204), in part by the Guangdong Basic and Applied Basic Research Foundation (Grant No.2023B1515120057), in part by the Education Bureau of Guangzhou.

5.Applications

Example: Make a city in Blender.

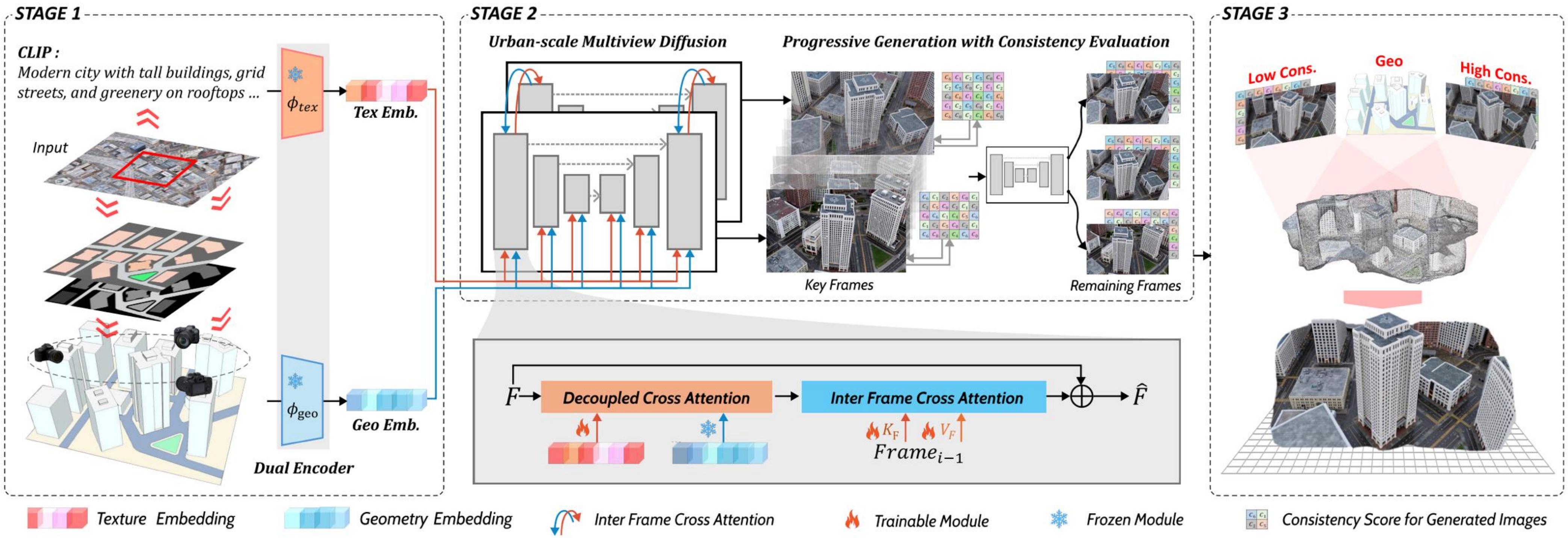
MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency

- **Motivation:**
  - Generating 3D cities supports autonomous driving simulations and model training.
  - However, traditional models struggle to generate large scenes with high consistency.
- **Achievements:**
  - Given satellite input, our method generates multi-view images with *city-level consistency*.
  - These images are then fed into a *robust reconstruction* pipeline to generate a 3D city.
  - Our approach achieves *diverse style* generation while maintaining *geometric consistency* across views.



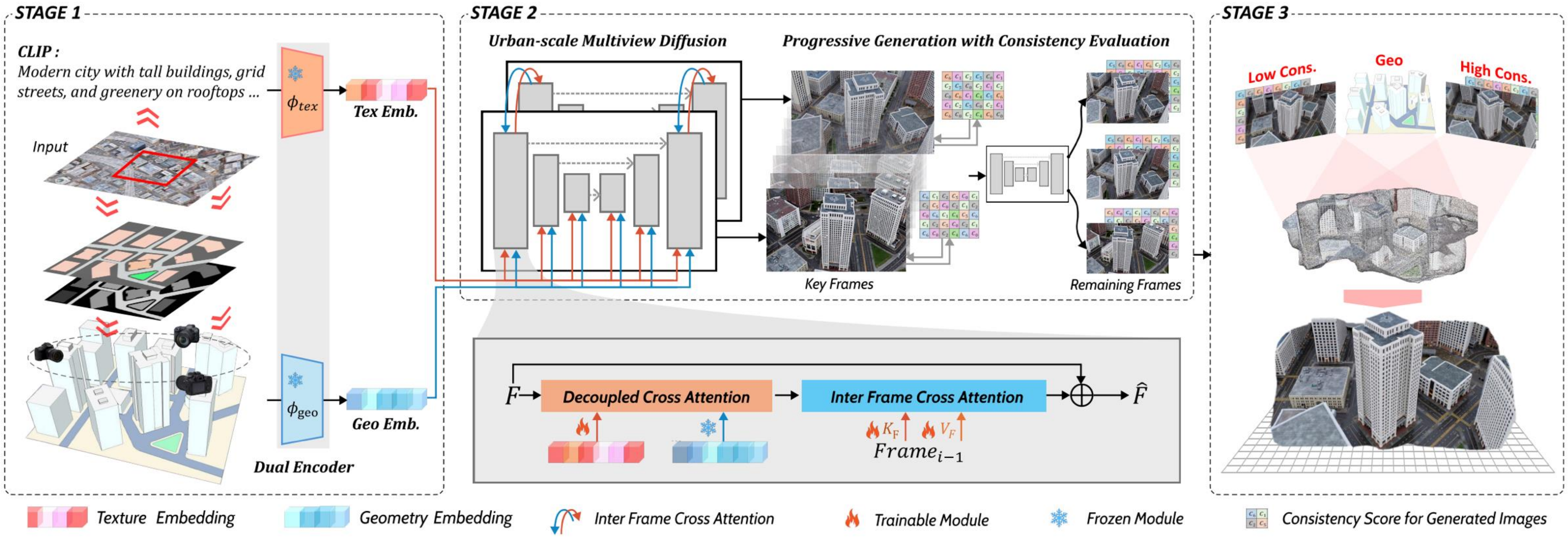
MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency

- **Key Contributions :**
  - We introduce **MagicCity**, a novel framework to generate photorealistic 3D cities from satellite imagery while maintaining scene-level geometric consistency.
  - We propose a city-scale multi-view diffusion model that generates 3D-consistent images by incorporating explicit geometric constraints.
  - We develop a robust 3D Gaussian Splatting strategy for synthesizing detailed 3D reconstructions from generated multi-view images.



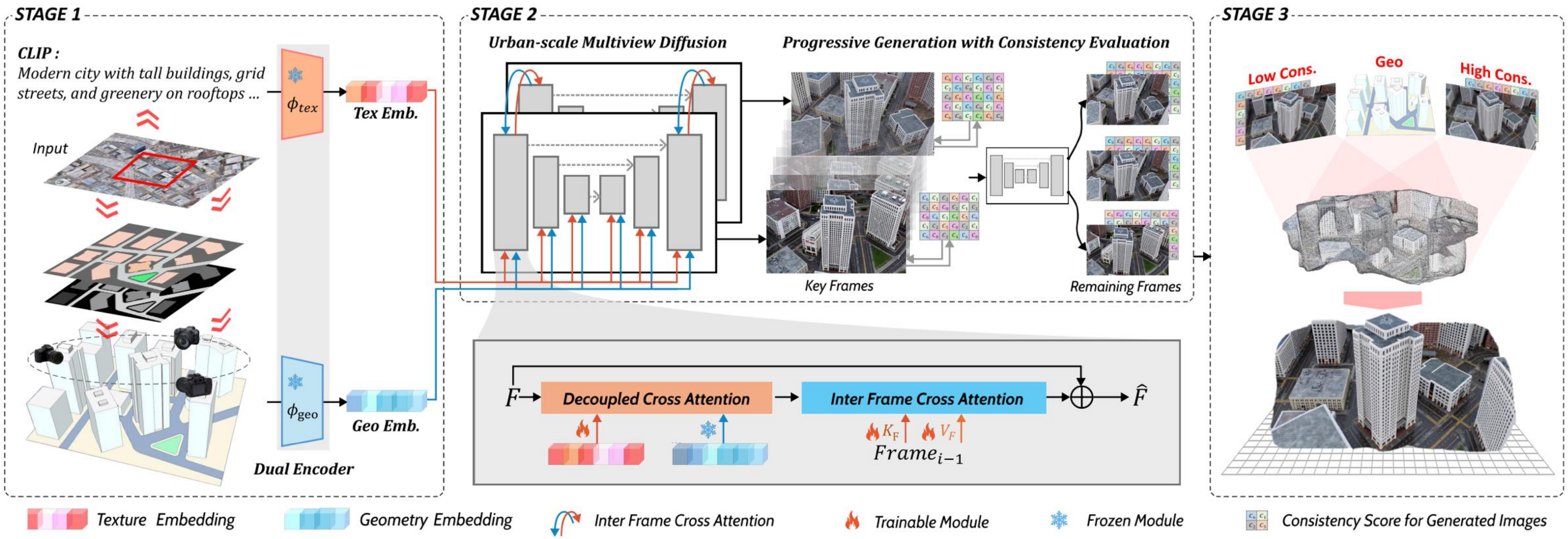
MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency

- ***Stage 1. Scene Initialization and Dual Encoding***
  - **Input:** Satellite images (x, y, r, g, b)
  - **Scene Initialization：** Segmentation+ Depth maps → 3D volume (x, y, z, r, g, b, s)
  - **Dual Encoder：**
    - **Texture**: CLIP → semantic tokens → texture features (f_{tex})
    - **Geometry**: 3D volume → multi-view renders → geometric features (f_{geo}^{i})



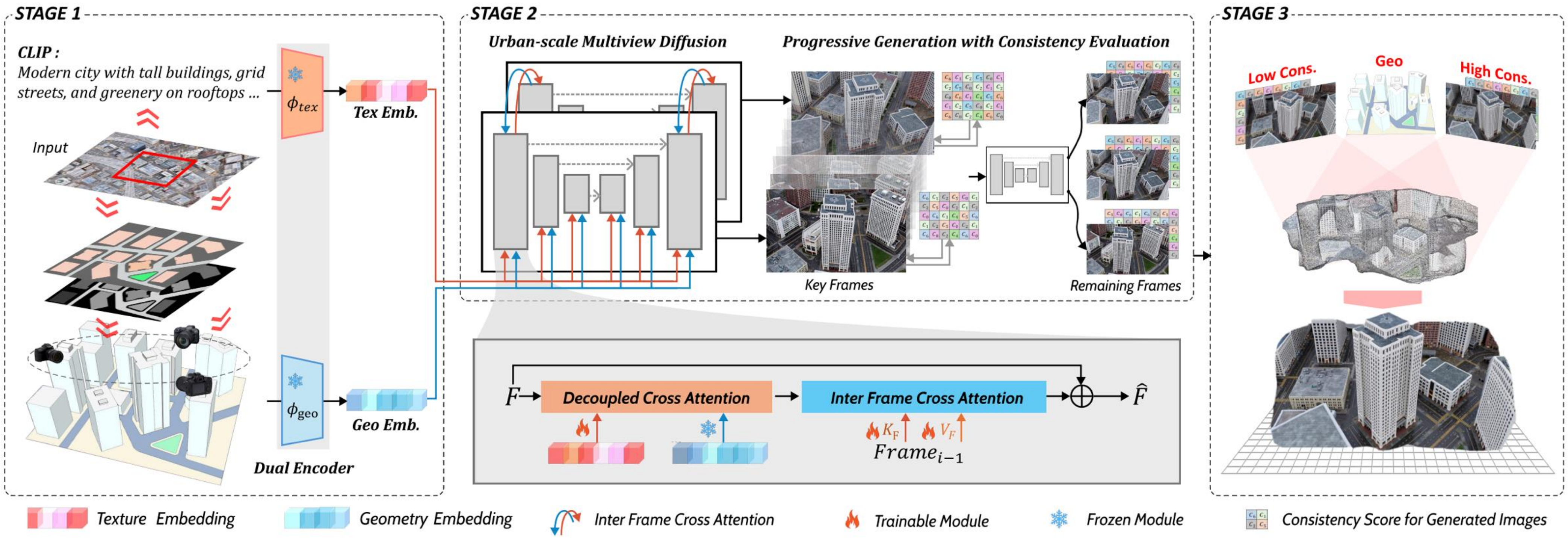
MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency

- ***Stage 2. City-scale Multi-view Generation with Consistency Evaluation***
  - **Goal**：Generate multi-view images with city-level consistency
  - (1) Dual Embedding Injection: $\tilde{F}_i = F_i + \mathcal{A}(F_i, f_{tex}) + \mathcal{A}(F_i, f^i_{geo})$
  - (2) Inter-Frame Cross-Attention: $M_i = \sum_{l \neq i} \mathrm{softmax}(W_q\tilde{F}_i \cdot W_k\tilde{F}_r^T) \cdot W_v\tilde{F}_r$
  - (3) Consistency Evaluation: $C^k_i = \frac{1}{|V_i|} \sum_{j \in V_i} \cos(f^k_i, f^j_i)$



MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency

- ***Stage 3. Consistency Score-guided 3D Reconstruction***
  - Challenge: Even SOTA video generation models lack pixel-level consistency → poor 3DGS
  - Reconstruct 3D cities guided by consistency scores
  - Prioritize colors from high-confidence pixels or images
  - Point Cloud Initialization: $c_p = \frac{\sum_k (C_i^k \cdot c_k)}{\sum_k C_i^k}$ Optimization: $L_p = L_{\text{render}}^p \cdot C_p$



MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency

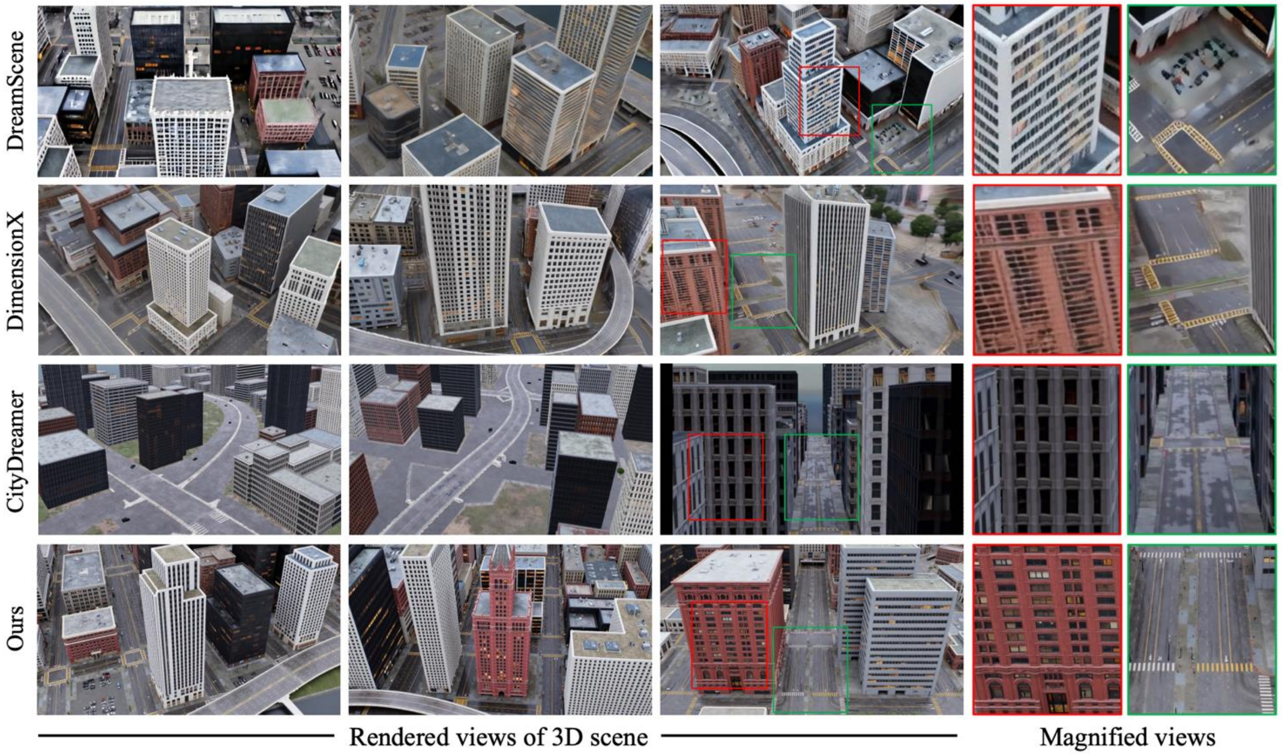
Results：

1. Qualitative Comparison

- Superior structure preservation compared to competitors；
- More realistic textures for buildings and ground details；
- Rich variety without synthetic artifacts；

2. Quantitative Evaluation

- Distribution Metrics: Lowest FID (86.096) and KID (0.087)
- Perceptual Quality: Best NIQE (4.553) and BRISQUE (28.018)
- Geometric Accuracy: Lowest Depth Error (0.137) and Camera Error (0.072)



Method	FID ↓	KID ↓	NIQE ↓	BRISQUE ↓	DE ↓	CE ↓
CityDreamer [38]	155.390	$0.251^{\pm 0.012}$	$8.632^{\pm 0.709}$	$86.773^{\pm 11.492}$	0.157	0.083
DimensionX [35]	126.890	$0.175^{\pm 0.006}$	$6.595^{\pm 0.425}$	$70.207^{\pm 7.157}$	-	-
DreamScene [17]	104.627	$0.125^{\pm 0.002}$	$6.018^{\pm 0.671}$	$30.311^{\pm 11.732}$	0.223	0.371
Ours	**86.096**	$\mathbf{0.087}^{\pm 0.001}$	$\mathbf{4.553}^{\pm 0.412}$	$\mathbf{28.018}^{\pm 5.634}$	**0.137**	**0.072**