Beyond human imagination: The art of creating prompt-driven 3D scenes with Generative AI

Giulio Federico
CNR-ISTI
Pisa, Italy
giulio.federico@isti.cnr.it

Fabio Carrara
CNR-ISTI
Pisa, Italy
fabio.carrara@isti.cnr.it

Giuseppe Amato
CNR-ISTI
Pisa, Italy
giuseppe.amato@isti.cnr.it

Marco Di Benedetto
CNR-ISTI
Pisa, Italy
marco.dibenedetto@isti.cnr.it

Is it possible to generate complex 3D scenes starting from a prompt?

Text prompt

```
A city with tall buildings
surrounded by trees...
```

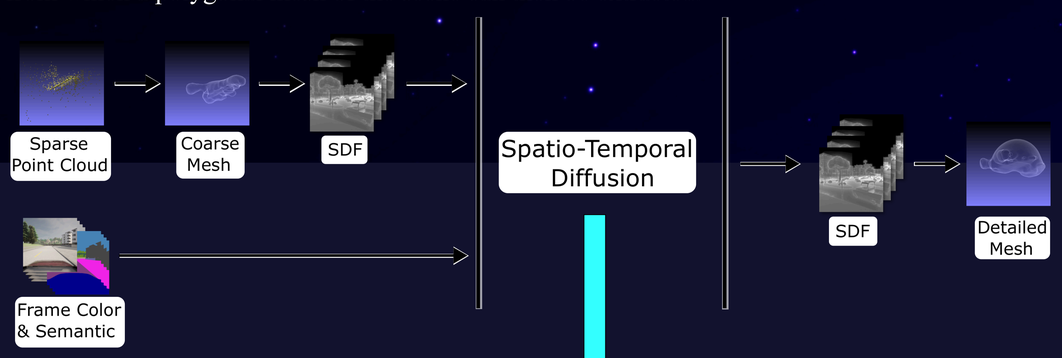
Images + Point Cloud prompt

Sketch prompt

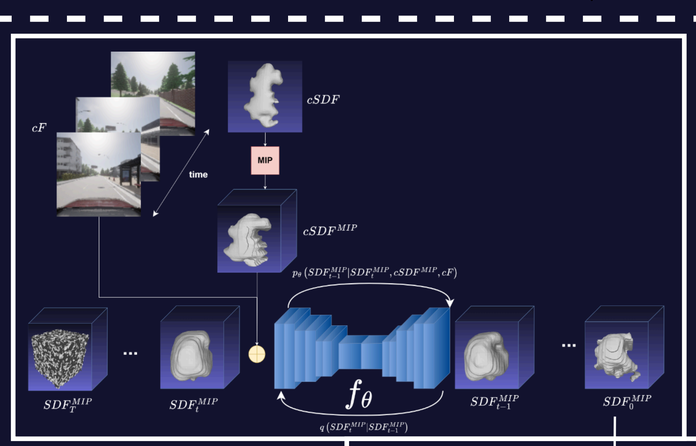
Brain prompt

1 What do we want to solve?

Reconstructing a large-scale real environment with limited resources is a complex challenge. Starting from few **color images** of a sequence, along with their corresponding **semantic maps** and **sparse spatial information** encoded as a **volumetric representation**, we proposed ***Spatio-Temporal 3D Reconstruction from Frame Sequences and Feature Points,** a **generative neural architecture** capable of generating a more detailed volumetric representation from which a polygonal mesh of the scene can then be extracted.



2 Our proposed Architecture



Stage 1

Generates a more detailed version of the coarse SDF ($cSDF^{MIP}$) by exploiting frames and their semantic maps (cF).

MIP is a new SDF-based representation that is more suitable for the task.

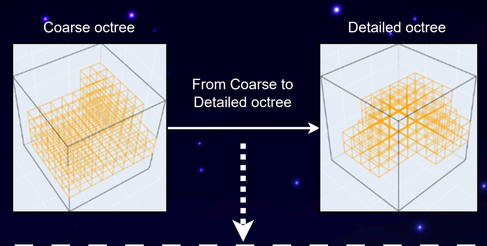
Stage 2

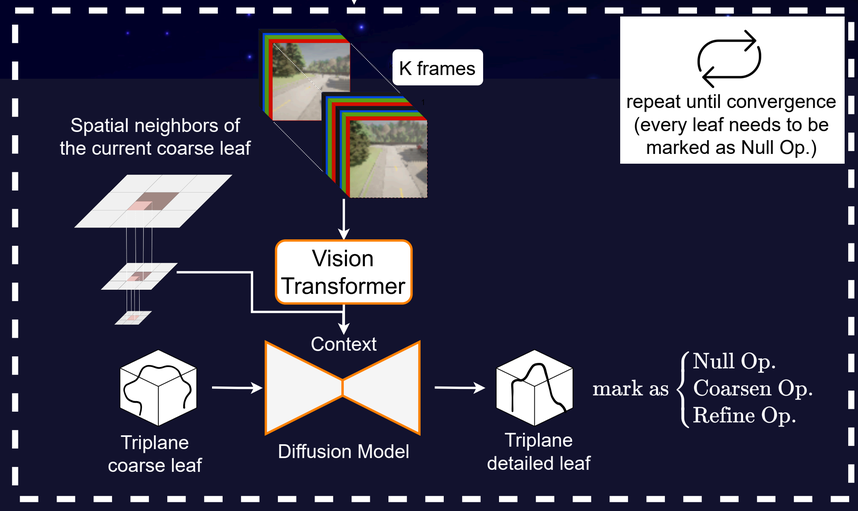
Convert MIP to SDF.

3 Our Results

4 Good results, but does not scale well with resolution

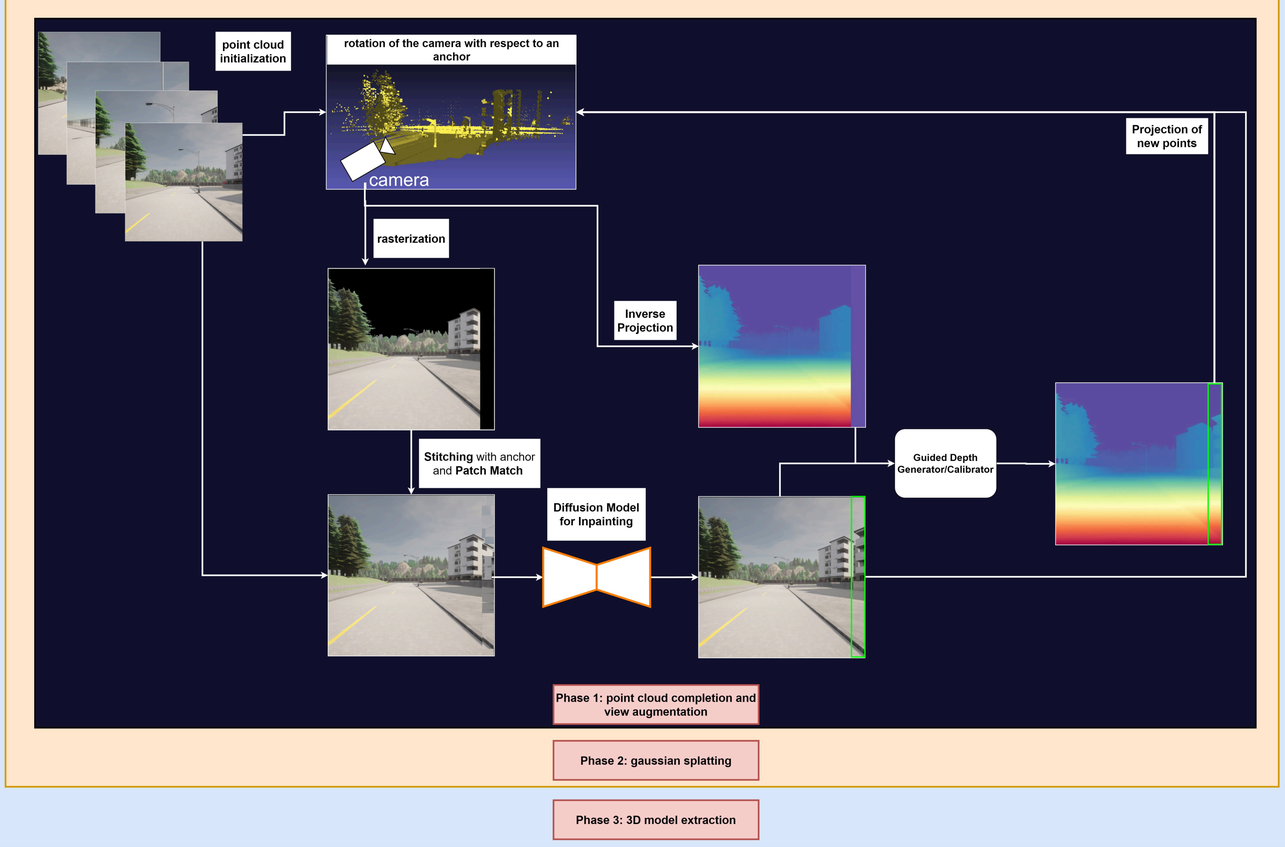
What if we did the generation on the octree instead of the entire SDF? **Neural-Clipmap** is an algorithm that iteratively transforms the coarse Octree into a detailed Octree **by operating only on individual leaves**.





The algorithm iterates in **two phases** until convergence: first, it generates a more detailed leaf from a coarse leaf using **contextual information**. Second, it prunes or increases the depth of the octree respectively for leaves marked Coarsen and Refine.

5 Good approach, but slow training and dataset preparation, leaves increase exponentially



* Giulio Federico, Fabio Carrara, Giuseppe Amato and Marco Di Benedetto. Spatio-Temporal 3D Reconstruction from Frame Sequences and Feature Points. In Video4IMX: 1st International Workshop on Video for Immersive Experiences, 2024.