# Aesthetic Camera Viewpoint Suggestion with 3D Aesthetic Field

Sheyang Tang, Armin Shafiee Sarvestani, Jialu Xu, Xiaoyu Xu, Zhou Wang

UNIVERSITY OF WATERLOO

香港城市大學
City University of Hong Kong

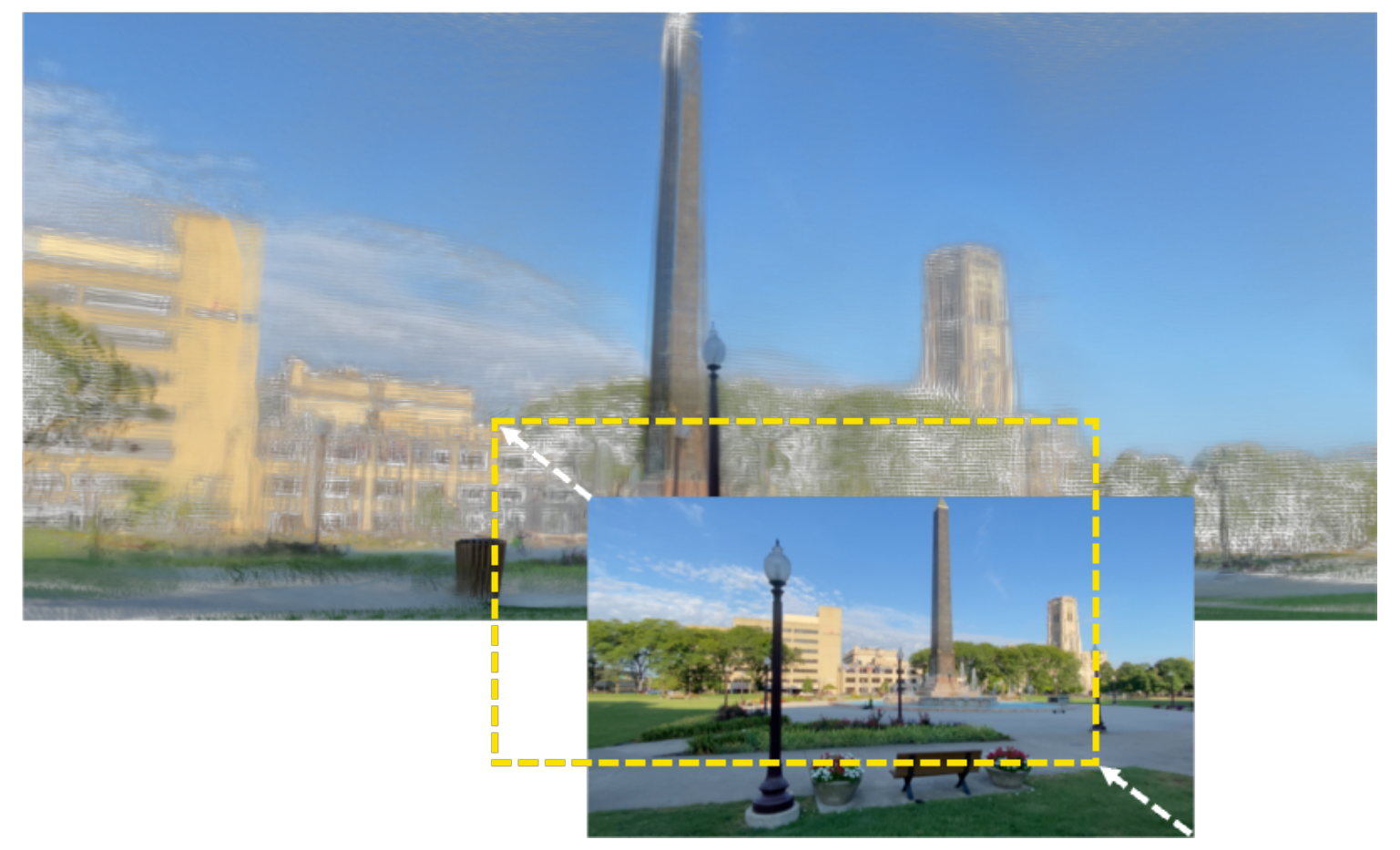
CVPR
JUNE 3-7, 2026
DENVER COLORADO

## Task

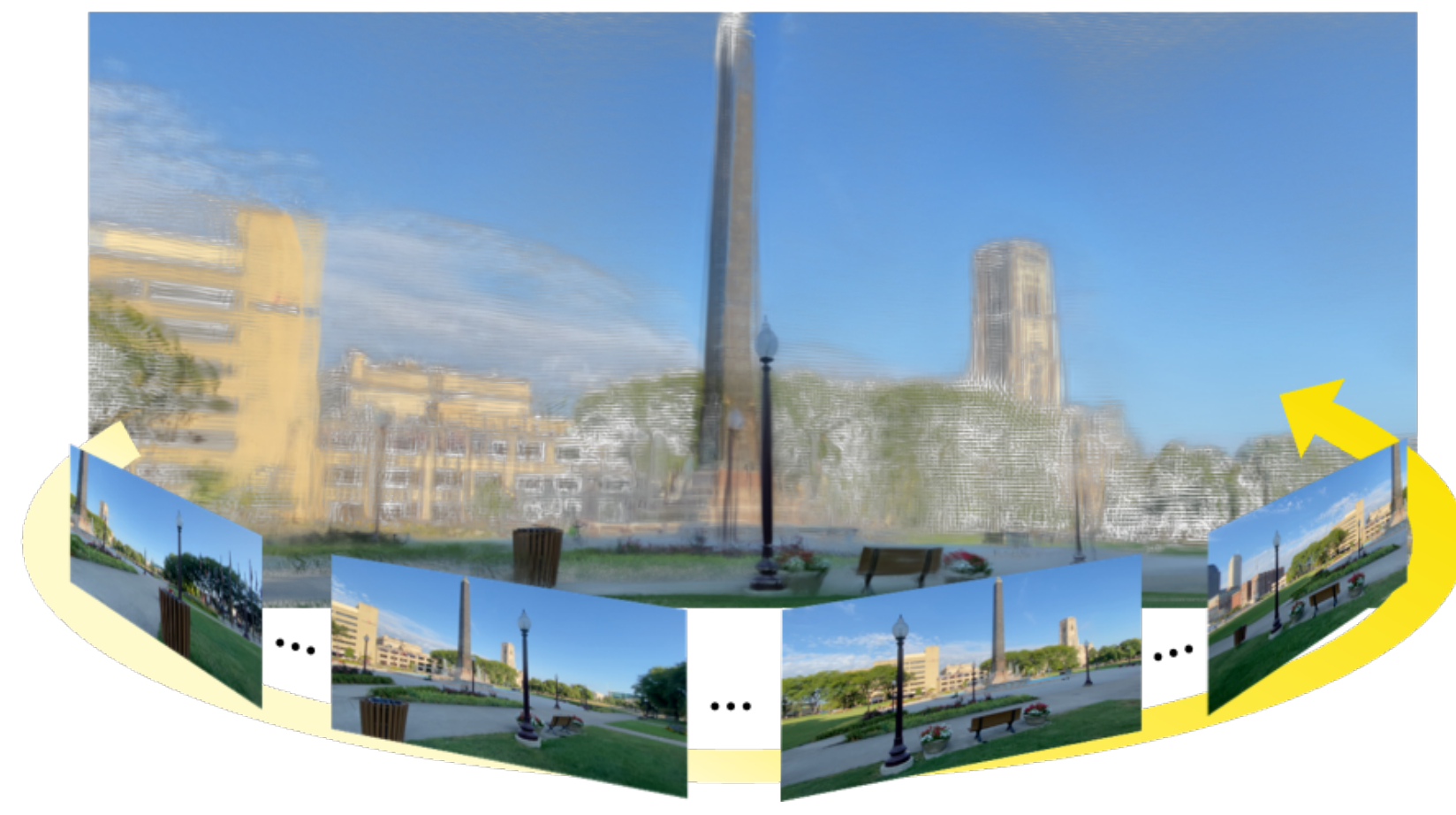
📷 How to find the camera pose that gives the best ***framing and composition*** of a ***3D scene***, *given* ***few observations***?

This can be useful for **photography**, **view selection/planning in VR/AR**, and **autonomous capture systems**.

❌ Existing approaches:



Single view adjustment
Pure 2D; No 3D understanding



3D exploration
RL search; Need dense captures

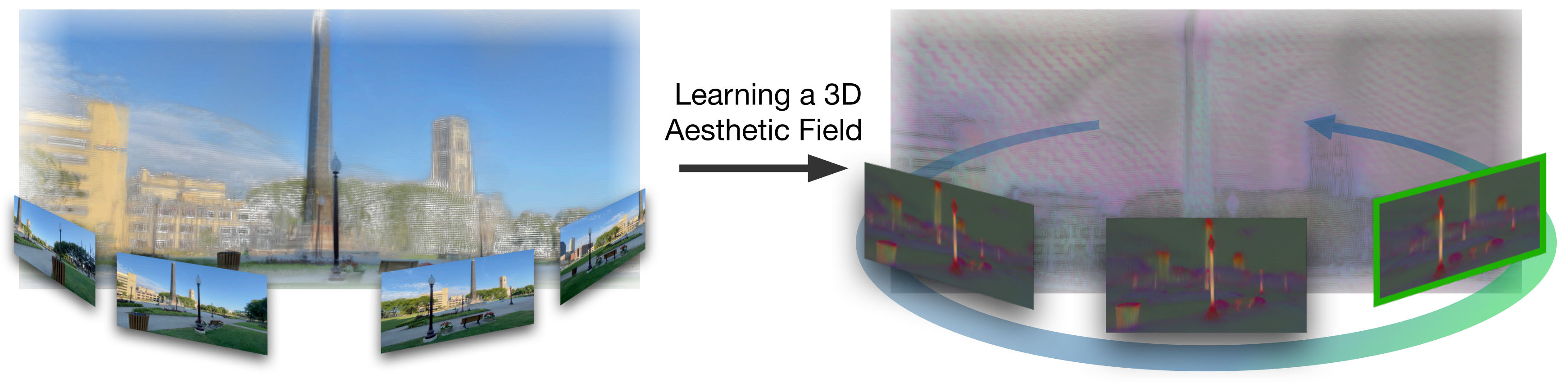
## Key Insights

🏞 Aesthetics is inherently ***3D-dependent***. As spatial layout, perspectives, object arrangement all change with viewpoints.

From only a few observations, humans can form a rough ***mental map*** of how aesthetic quality changes across viewpoints.

## Solution

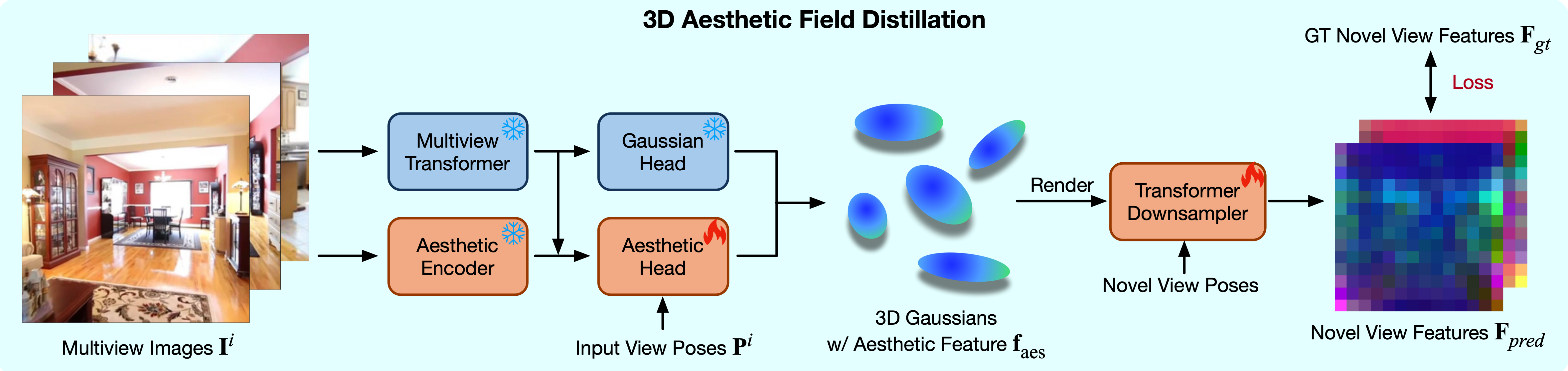
A 3D aesthetic field for fast viewpoint search with sparse captures.


Learning a 3D Aesthetic Field

Sparse Scene Captures

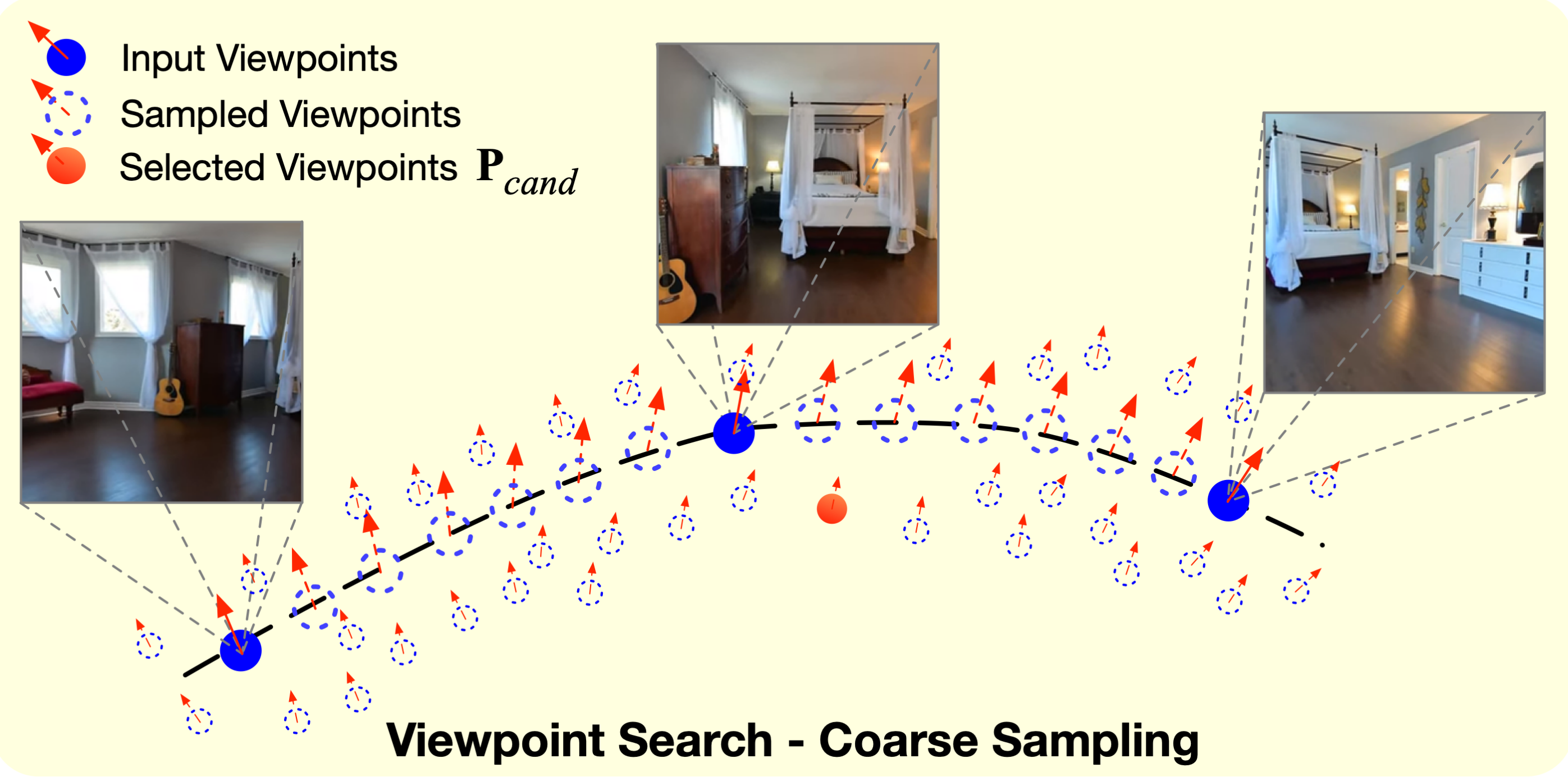
3D Aesthetic Field and Viewpoint Search

## Learning a 3D Aesthetic Field

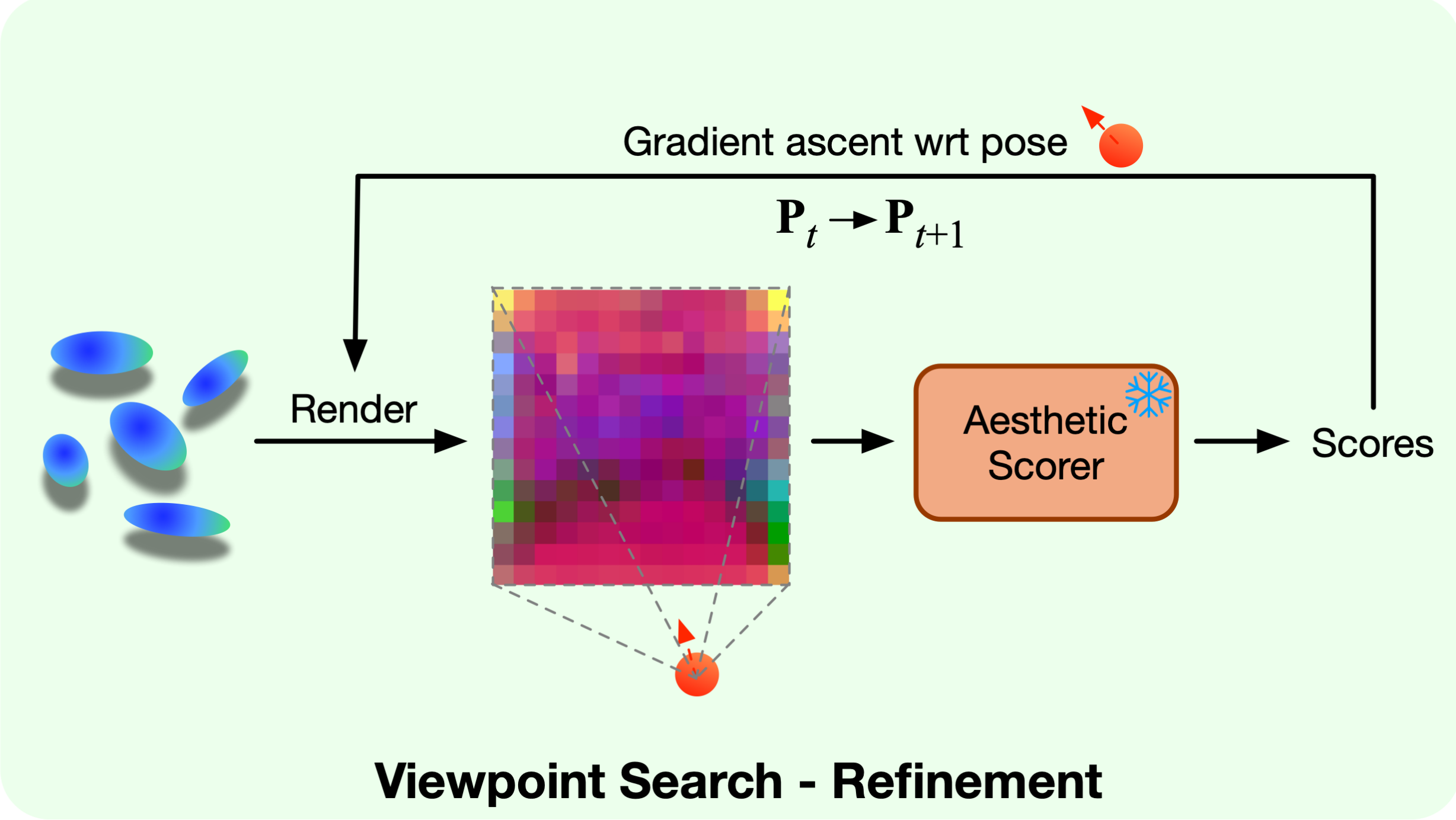


**3D Aesthetic Field Distillation**

Multiview Images $\mathbf{I}^i$
Multiview Transformer
Aesthetic Encoder
Gaussian Head
Aesthetic Head
Input View Poses $\mathbf{P}^i$
3D Gaussians w/ Aesthetic Feature $\mathbf{f}_{aes}$
Render
Transformer Downsampler
Novel View Poses
GT Novel View Features $\mathbf{F}_{gt}$
Loss
Novel View Features $\mathbf{F}_{pred}$


Input Viewpoints
Sampled Viewpoints
Selected Viewpoints $\mathbf{P}_{cand}$

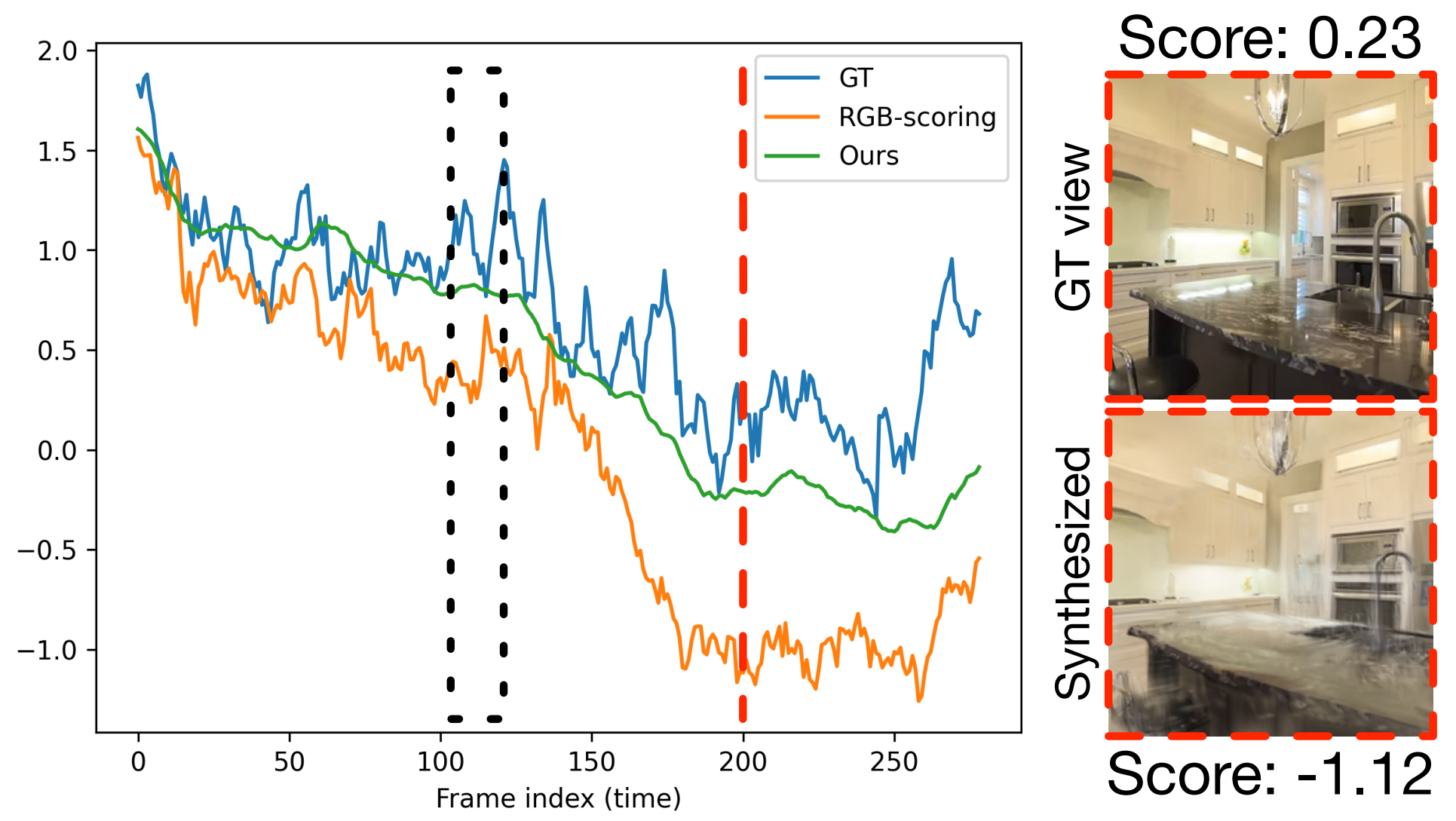
**Viewpoint Search - Coarse Sampling**


Gradient ascent wrt pose
$\mathbf{P}_t \rightarrow \mathbf{P}_{t+1}$
Render
Aesthetic Scorer
Scores

**Viewpoint Search - Refinement**

We distill aesthetic features from an image aesthetic model into a ***feed-forward 3DGS*** backbone. Novel-view aesthetics can be predicted by ***rendering and decoding*** novel-view aesthetic features. Visually pleasing views are found by an efficient two-stage viewpoint search pipeline.

## Instability of Direct RGB Scoring



**Issue 1.** Rendering artifacts can bias prediction and misguide optimization.

**Issue 2.** Aesthetic models are sensitive to pixel-level variations: adjacent GT views with nearly identical content receive fluctuated aesthetic scores.

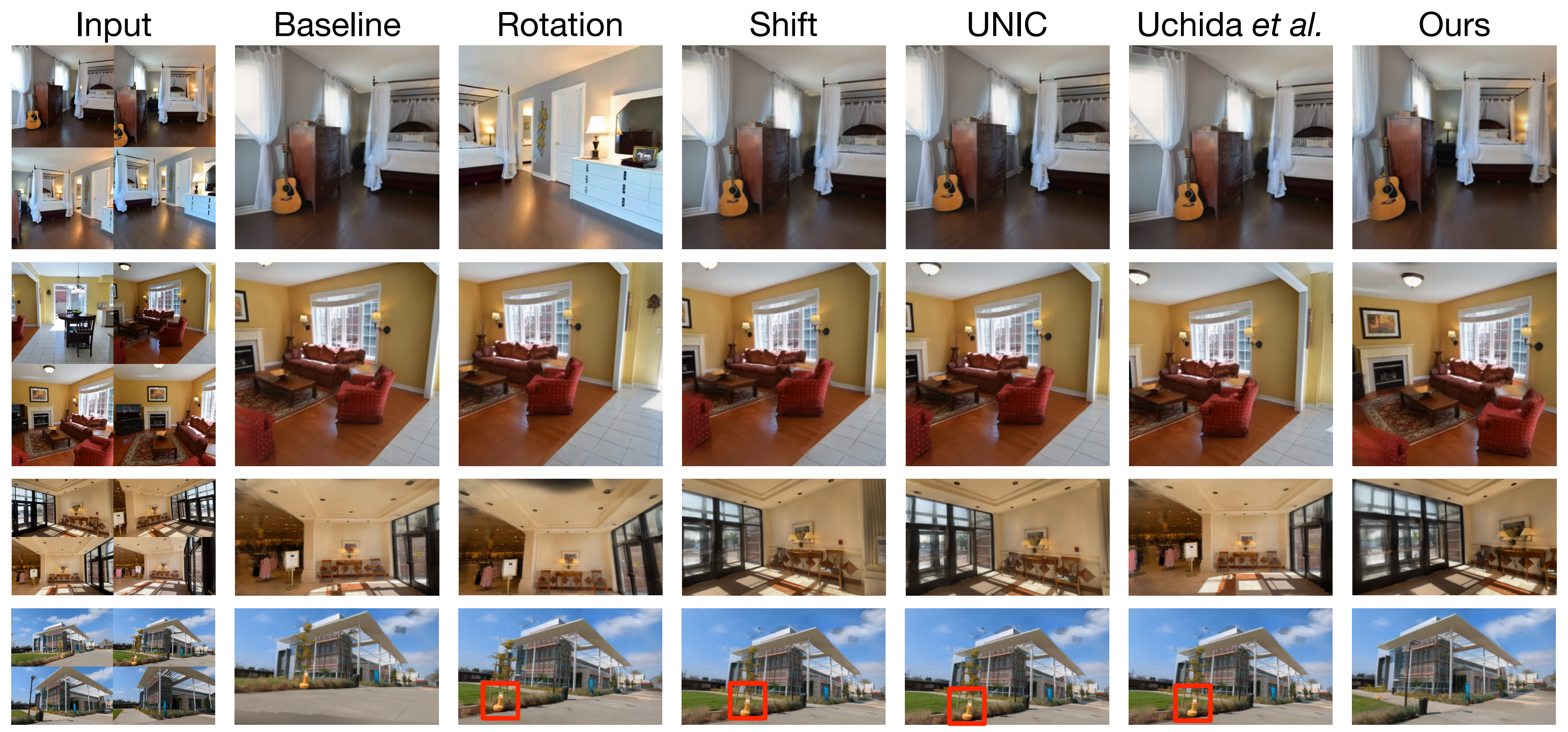
## Experiments

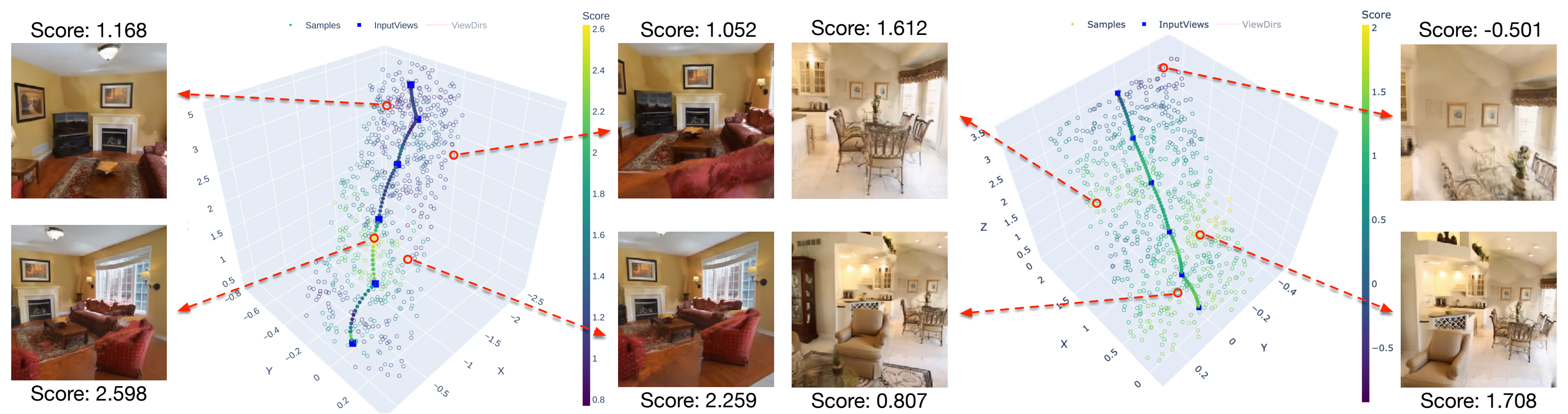
*RE10k* [40]

| Methods | 2 Input Views VEN↑ | 2 Input Views SAMPNet↑ | 4 Input Views VEN↑ | 4 Input Views SAMPNet↑ | 6 Input Views VEN↑ | 6 Input Views SAMPNet↑ |
|---|---|---|---|---|---|---|
| Baseline | 1.48 | 2.29 | 1.79 | 2.26 | 2.01 | 2.29 |
| In-plane Shift* | 1.52 | 2.31 | 1.81 | 2.36 | 2.10 | 2.41 |
| Rotation* | 1.78 | 2.38 | 1.95 | 2.42 | 2.13 | 2.45 |
| UNIC [14]† | 1.15 | 2.17 | 1.61 | 2.33 | 1.82 | 2.35 |
| Uchida *et al.* [24]† | 1.58 | 2.32 | 1.89 | 2.37 | 2.13 | 2.42 |
| **Ours** | **1.89** | **2.40** | **2.03** | **2.45** | **2.20** | **2.49** |

*DL3DV* [13]

| Methods | 2 Input Views VEN↑ | 2 Input Views SAMPNet↑ | 4 Input Views VEN↑ | 4 Input Views SAMPNet↑ | 6 Input Views VEN↑ | 6 Input Views SAMPNet↑ |
|---|---|---|---|---|---|---|
| Baseline | 2.08 | 2.23 | 2.31 | 2.26 | 2.47 | 2.30 |
| In-plane Shift* | 2.16 | 2.28 | 2.45 | 2.29 | 2.59 | 2.33 |
| Rotation* | 2.52 | 2.30 | 2.67 | 2.32 | 2.85 | 2.37 |
| UNIC [14]† | 2.07 | 2.21 | 2.35 | 2.27 | 2.48 | 2.30 |
| Uchida *et al.* [24]† | 2.34 | 2.25 | 2.68 | 2.32 | 2.81 | 2.36 |
| **Ours** | **2.56** | **2.33** | **2.76** | **2.36** | **2.91** | **2.41** |


Input
Baseline
Rotation
Shift
UNIC
Uchida et al.
Ours

## Visualizations