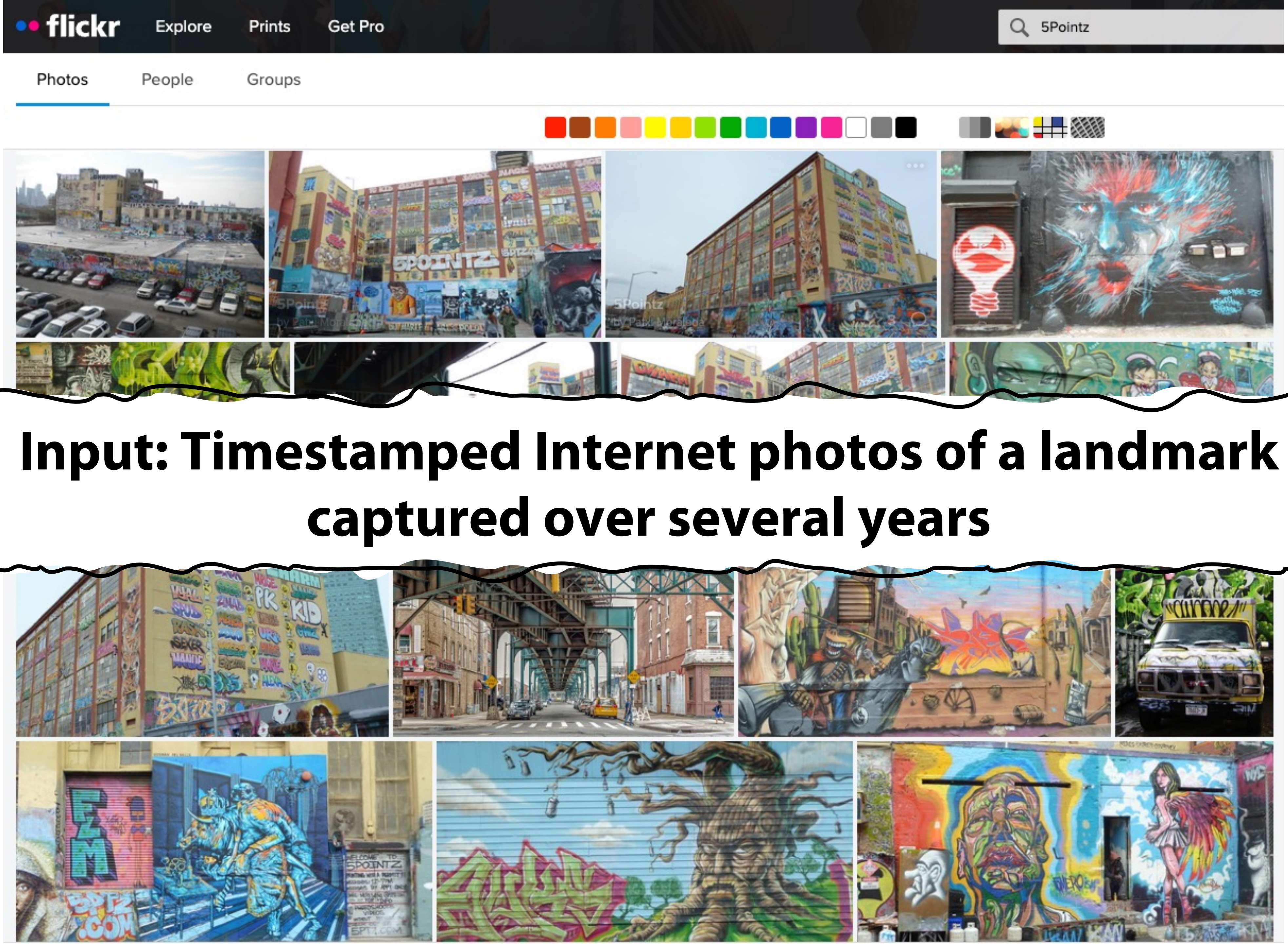
Poster Session:
THU-PM-014

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Neural Scene Chronology

Haotong Lin[1], Qianqian Wang[2], Ruojin Cai[2], Sida Peng[1],
Hadar Averbuch-Elor[3], Xiaowei Zhou[1] and Noah Snavely[2]

1. Zhejiang University, 2. Cornell University, 3. Tel Aviv University



Input: Timestamped Internet photos of a landmark captured over several years

Output: 4D reconstruction with controllable viewpoint, time, and illumination

2013–10–31

Problem Motivation

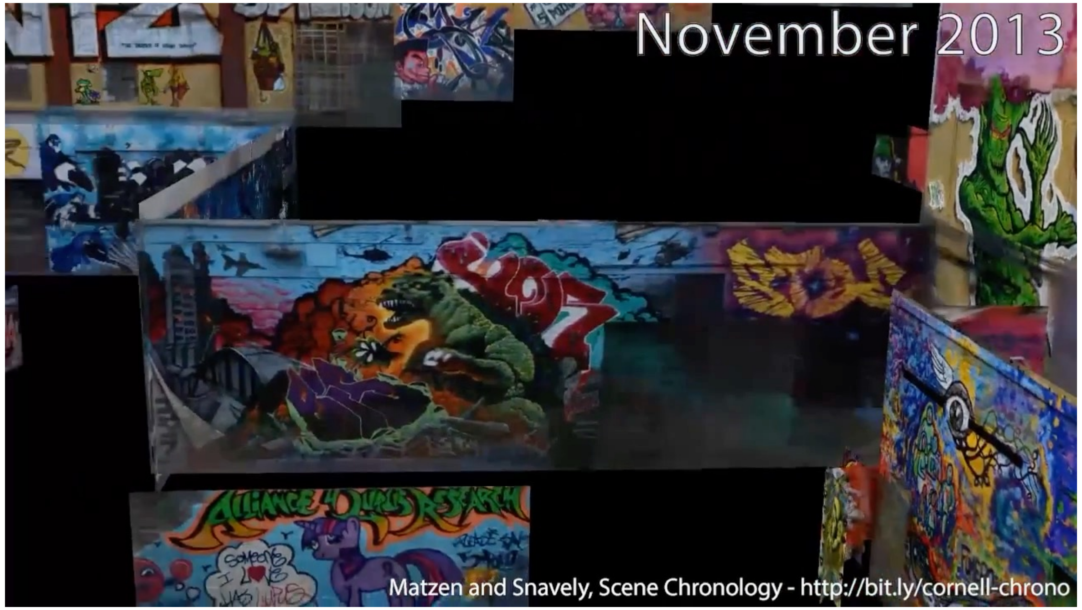
Chronology reconstruction

- Landmarks evolve over time.
- A 3D reconstruction can only capture a certain moment of the landmark.
- We need **Chronology Reconstruction!**

Times Square, February, 2011

Times Square, December, 2011

Previous methods: Scene Chronology



5PointZ

Times Square

Scene Chronology [Matzen & Snavely, ECCV 2014] only reconstructs planes,
resulting in limited photo-realism.

Previous methods: NeRF-W, HaNeRF

Trevi Fountain (Nearly constant over time)

5Pointz (The underlying apperance changes significantly over time)

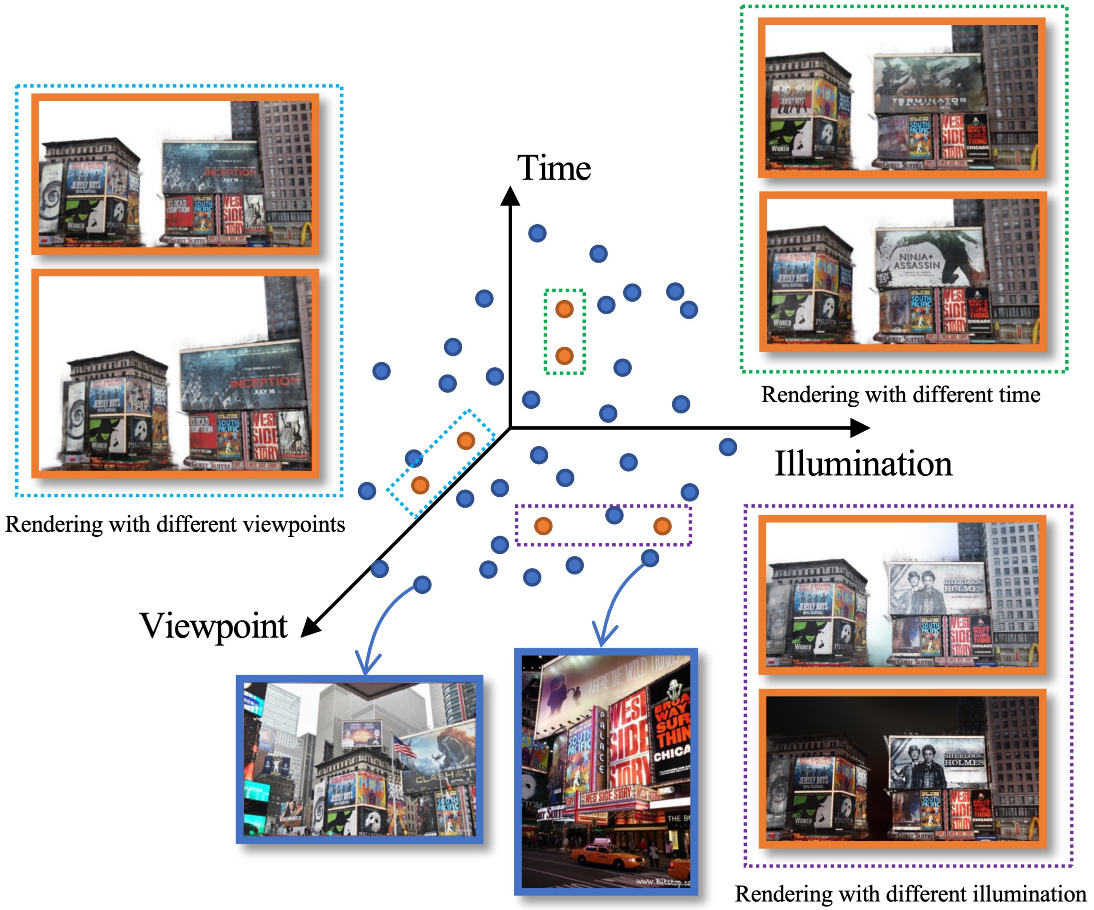
NeRF-W [Martin-Brualla et al., CVPR 2021] and HaNeRF [Chen et al., CVPR 2022] only handle static scenes (e.g., Trevi Fountain) and **cannot reconstruct a scene (e.g., 5Pointz) with significant changes to the underlying appearance over years.**

Proposed Method

Key challenge

- This problem is highly challenging, as every photo entangles the **viewpoint, scene content (time) and illumination.**

- We need to **decompose** these factors, achieving independent control.



4D reconstruction from Internet photos

- We represent such a time-varying landmark as color $\boldsymbol{c}$ and density σ fields.

$$\boldsymbol{c}, \sigma = \mathrm{F}(\boldsymbol{x}, \boldsymbol{d}, \mathrm{t_i}, \mathrm{l_i})$$

Technical challenges

- Fitting above model to a set of time-stamped images **underfits** scene-level temporal changes, blending different scene appearances together.

Technical challenges

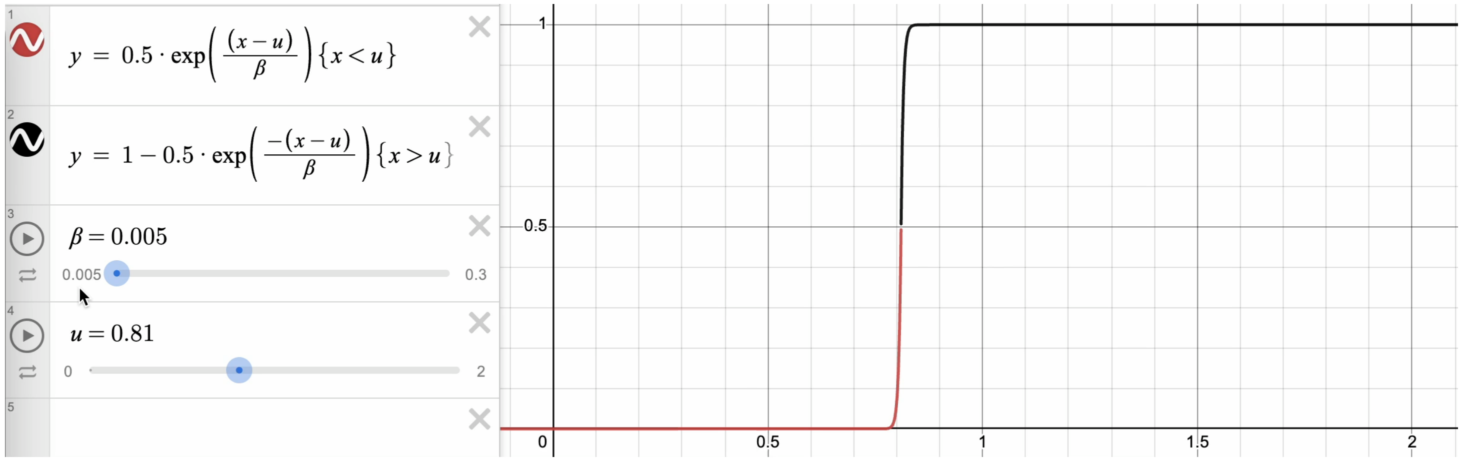
- Fitting above model to a set of time-stamped images **underfits** scene-level temporal changes, blending different scene appearances together.

- In contrast, **applying positional encoding to the time input overfits the temporal signal.**

Step function encoding

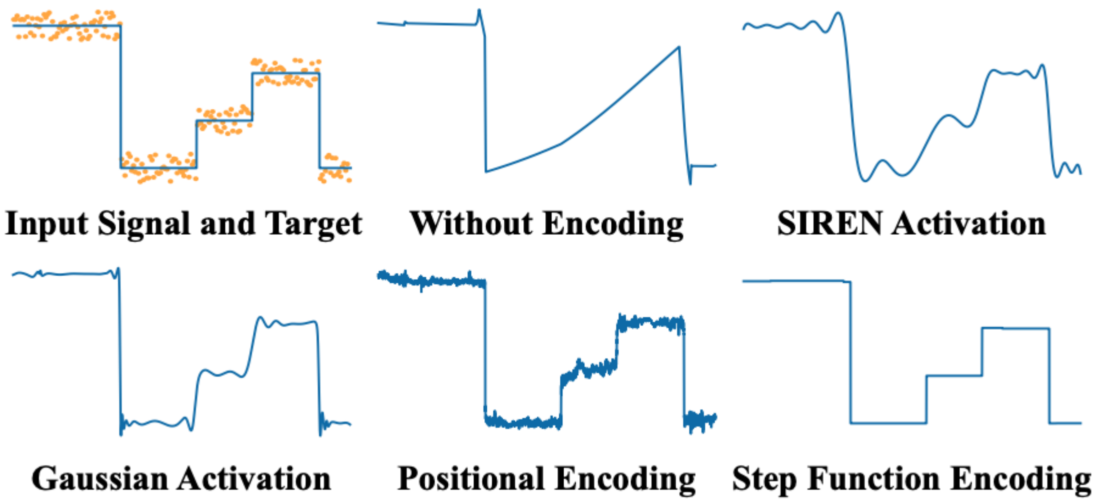
- Changes in the underlying content of urban scenes often **happen abruptly**, and this content typically **remains constant for a period** after these changes. We introduce the step function encoding.

$$\bar{h}(t) = \begin{cases} \frac{1}{2}\exp(\frac{t-u}{\beta}) & \text{if } \; t \leq u \\ 1 - \frac{1}{2}\exp(\frac{-(t-u))}{\beta}) & \text{if } \; t > u \end{cases}$$



Step function encoding

Formulation: Recovering a noiseless piecewise signal from a noisy signal.



Experimental Results

2013-10-31

Controlling the illumination using a reference image

Fixed viewpoint, changing Illumination

Times Square

Controlling the illumination using a reference image

Reference image

Fixed time, rendering with the illumination of the reference image

The Metropolitan Museum of Art

2010–06–30

Thank you for watching!

Code and Data: https://zju3dv.github.io/neusc