Preface: A Data-driven Volumetric Prior for Few-shot Ultra High-resolution Face Synthesis

Marcel C. Bühler, Kripasindhu Sarkar, Tanmay Shah, Gengyan Li, Daoye Wang, Leonhard Helminger, Sergio Orts-Escolano, Dmitry Lagun, Otmar Hilliges, Thabo Beeler, Abhimitra Meka
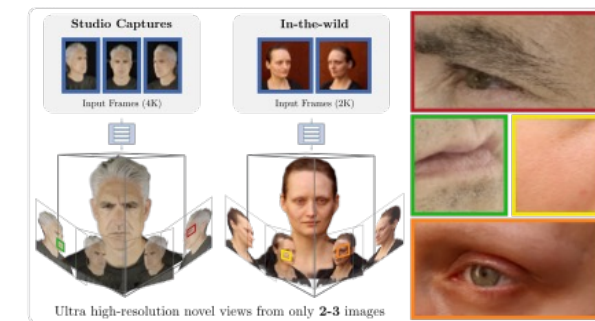
aitlab
ETH zürich

Motivation

Ultra high-resolution novel view synthesis of faces from **sparse** inputs is a highly challenging problem in 3D computer vision. With a strong **prior**, a Neural Radiance Field can be trained **on very few input images**.

Input — W/o Prior — With — Training Result — Novel Views
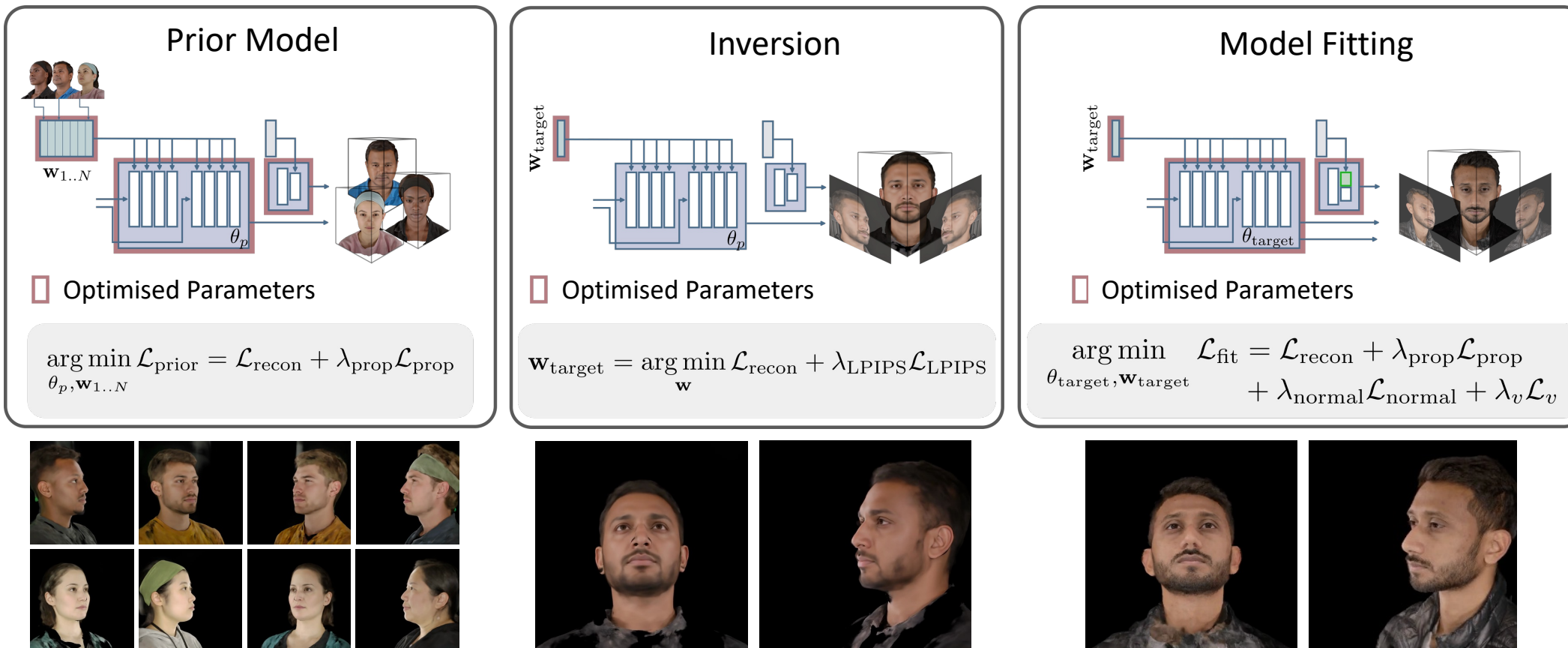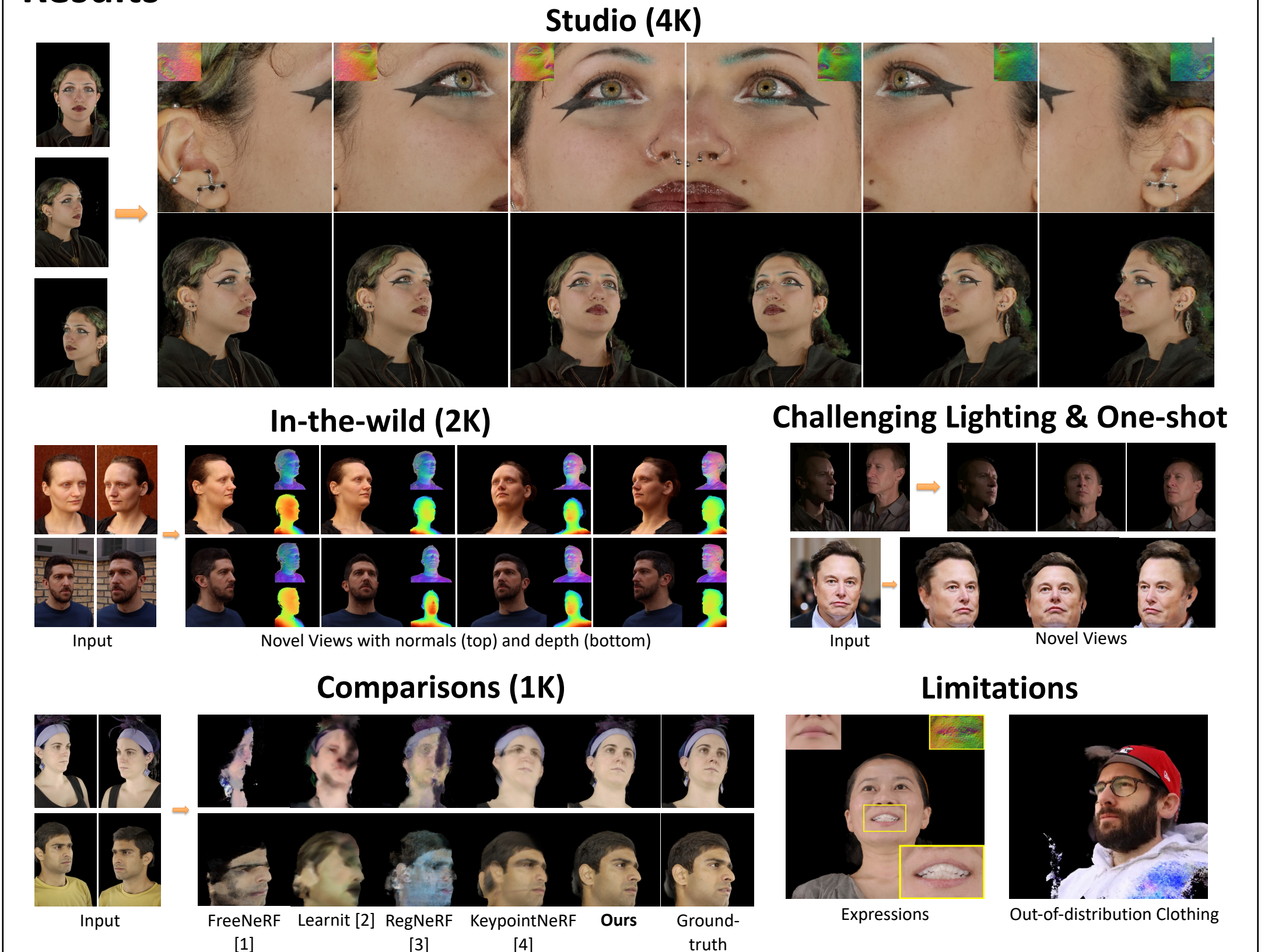
Contributions

1. A **prior** model for faces that can be finetuned to generate a **high-quality** volumetric 3D representation of a target identity from **sparse inputs**.
2. **Ultra high-resolution** 3D consistent **view-synthesis**.
3. Generalisation to **in-the-wild** indoor and outdoor captures, including challenging lighting conditions.

Ultra high-resolution novel views from only **2-3** images

Method

Prior Model

□ Optimised Parameters

$$\underset{\theta_p, \mathbf{w}_{1..N}}{\arg\min} \mathcal{L}_{\text{prior}} = \mathcal{L}_{\text{recon}} + \lambda_{\text{prop}}\mathcal{L}_{\text{prop}}$$

Inversion

□ Optimised Parameters

$$\mathbf{w}_{\text{target}} = \underset{\mathbf{w}}{\arg\min} \mathcal{L}_{\text{recon}} + \lambda_{\text{LPIPS}}\mathcal{L}_{\text{LPIPS}}$$

Model Fitting

□ Optimised Parameters

$$\underset{\theta_{\text{target}}, \mathbf{w}_{\text{target}}}{\arg\min} \quad \mathcal{L}_{\text{fit}} = \mathcal{L}_{\text{recon}} + \lambda_{\text{prop}}\mathcal{L}_{\text{prop}} + \lambda_{\text{normal}}\mathcal{L}_{\text{normal}} + \lambda_v \mathcal{L}_v$$

Results

Studio (4K)

In-the-wild (2K)

Input — Novel Views with normals (top) and depth (bottom)

Challenging Lighting & One-shot

Input — Novel Views

Comparisons (1K)

Input — FreeNeRF [1] — Learnit [2] — RegNeRF [3] — KeypointNeRF [4] — **Ours** — Ground-truth

Limitations

Expressions — Out-of-distribution Clothing

References
[1] Freenerf: Improving few-shot neural rendering with free frequency regularization, CVPR 2023. [2] Learned initializations for optimizing coordinate-based neural representations, CVPR 2021. [3] Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs, CVPR 2022. [4] KeypointNeRF: Generalizing image-based volumetric avatars using relative spatial encoding of keypoints, ECCV 2022.