InstantAvatar: Efficient 3D Head Reconstruction via Surface Rendering

Antonio Canela Pol Caselles Ibrar Malik Eduard Ramon Jaime Garcia
Jordi Sanchez-Riera Gil Triginer Francesc Moreno-Noguer

crisalixsa.github.io/instant-avatar

Introduction

Optimization based models
Representations

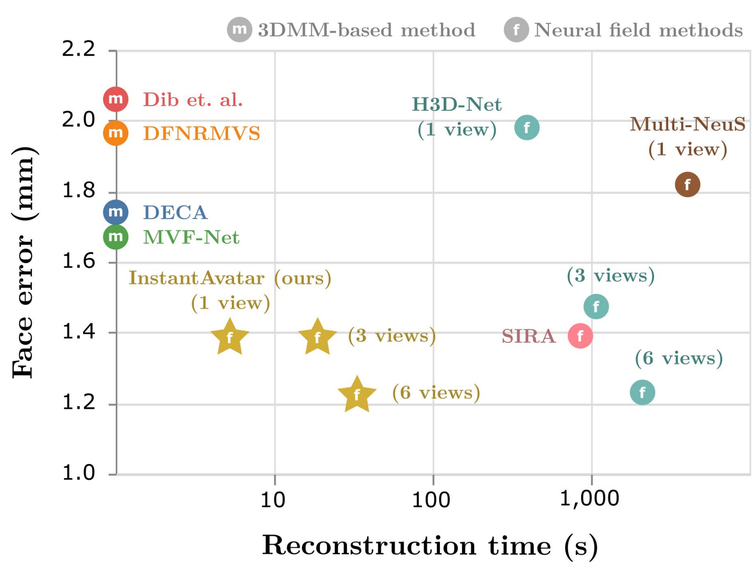
Voxelgrid | **Octree** | **Point cloud** | **Mesh** | **SDF**

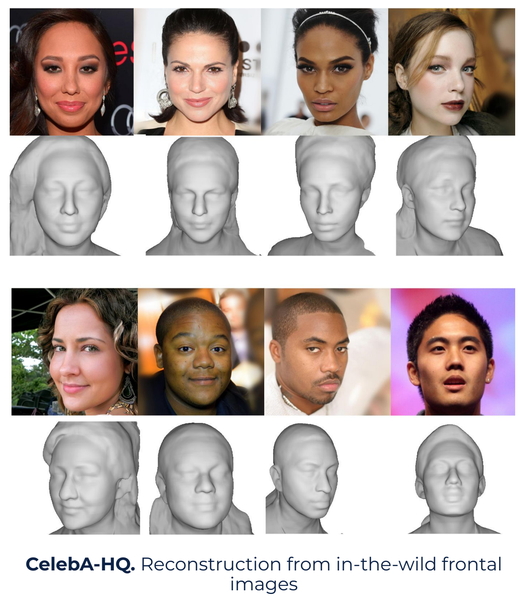
Optimization based models
Rendering

Method	Computational cost	Rendering stability	Use case
Surface rendering	**Lower**	**Lower**	**Objects**
Volumetric rendering	**Higher**	**Higher**	**Scenes**
Surface + Volumetric rendering	**Higher**	**Higher**	**Objects / Scenes**

Contributions

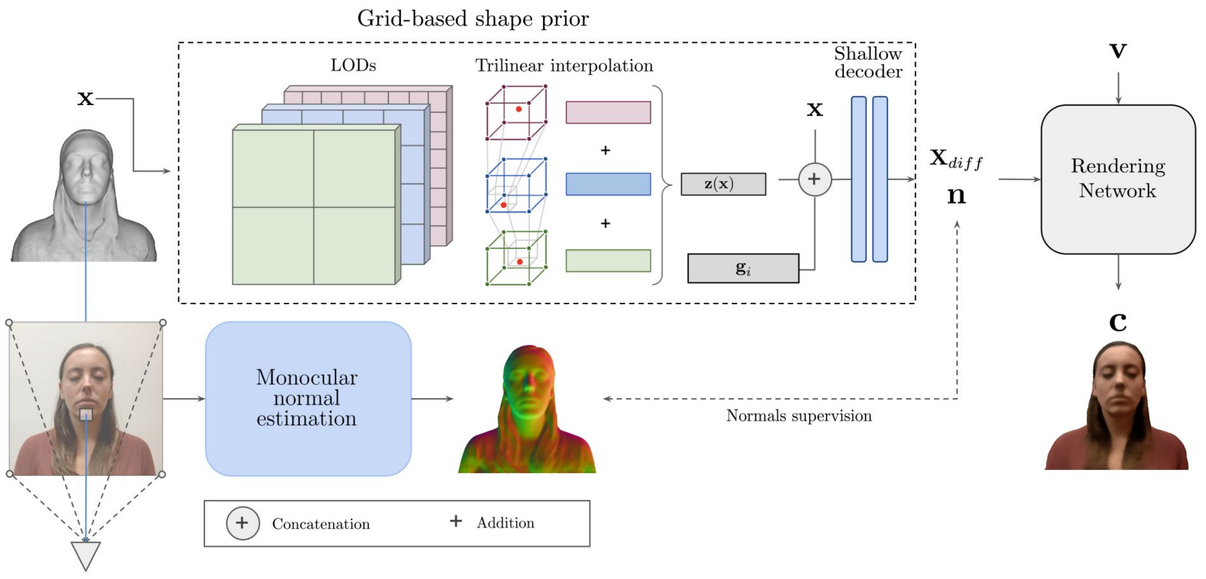
- We introduce, for the first time, **a framework that combines a grid-based architecture with a surface rendering method** that yields to fast and accurate 3D reconstructions **from one or few input images**.
- We leverage a statistical **prior**, obtained with thousands of 3D head models, to **guide network convergence** and achieve a reconstruction accuracy on a par with state of the art methods, but with **~100× speed-up**.
- We provide **optimal training techniques** for grid-based structures combined with surface rendering methods exhibited through a **variety of evaluated datasets.**.





CelebA-HQ. Reconstruction from in-the-wild frontal images

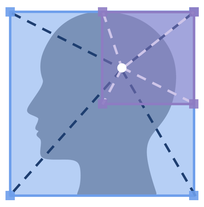
Method



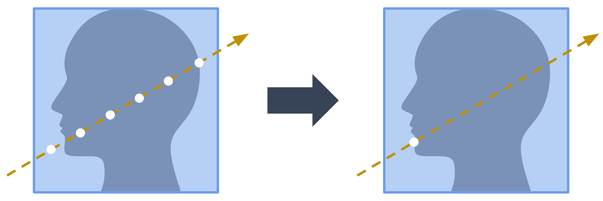
Key Insights

Speed

Grids improve convergence speed



Surface rendering has less computational cost than volumetric rendering



Stability

These speed upgrades **produce instability**, which we diminish by:

- **Statistical shape prior**
- **Progressive LODs scheduling**
- **Supervision with normals**

Ablation

	1 view	3 views	6 views
	face ↓	face ↓	face ↓
(A) Ours w/o progressive PE masking	1.77	1.48	1.33
(B) Ours w/o progressive LODs schedule	1.70	1.54	1.35
(C) **Ours**	**1.50**	**1.39**	**1.22**

Shape prior training schedules ablation.
Error[1] in millimeters.

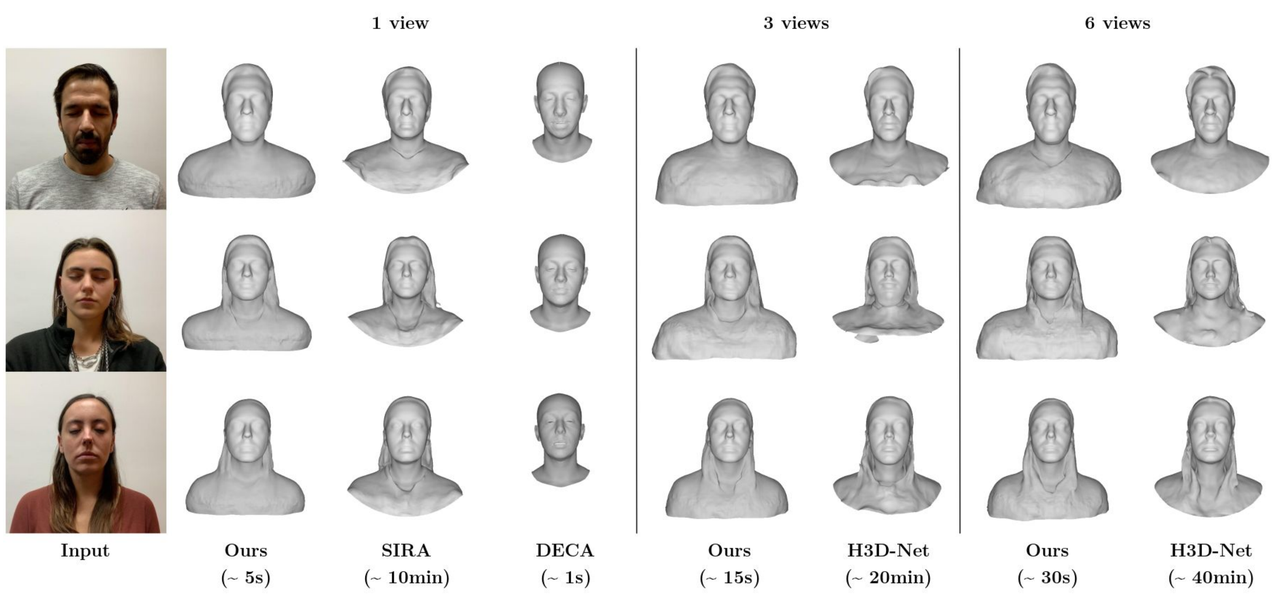
(1) We use the unidirectional Chamfer distance from the defined regions on the ground truth to the predicted reconstructions

Comparisons

	1 view					3 views				6 views		
	3DFAW	3DFAW HR	H3DS			3DFAW	H3DS			H3DS		
	face ↓	face ↓	face ↓	head ↓	time ↓	face ↓	face ↓	head ↓	time ↓	face ↓	head ↓	time ↓
Dib et al.	1.99	2.16	2.06	-	∼ **1s**							
DECA	1.57	**1.43**	1.74	22.22	∼ **1s**							
Multi-NeuS	-	-	1.81	13.92	∼ 1h							
H3D-Net	1.63	1.47	1.97	14.36	∼ 10min							
SIRA	**1.36**	1.51	**1.39**	14.43	∼ 10min							
Ours	1.58	1.69	1.50	**12.43**	∼ **5s**							
MVF-Net						1.61	1.67	-	∼ **1s**	-	-	-
DFNRMVS						1.74	1.96	-	∼ **1s**	-	-	-
H3D-Net						**1.32**	1.47	11.13	∼ 20min	1.24	**6.24**	∼ 40min
Ours						**1.32**	**1.39**	**9.71**	∼ **15s**	**1.22**	8.03	∼ **30s**

3D reconstruction method comparison (single-view).
Error[1] in millimeters. Time in seconds.

3D reconstruction method comparison (multi-view).
Error[1] in millimeters. Time in seconds.



References

[1] Dib et al. - ICCV 2021
[2] DECA: Feng et al. - TOG 2021
[3] H3D-Net: Ramon et al. - ICCV 2021
[4] SIRA: Caselles et al. - WACV 2023
[5] MVF-Net: Wu et al. - CVPR 2019
[6] DFNRMVS: Bai et al. - CVPR 2020
[7] Multi-NeuS: Burkov et al. - IEEE Access 2023