VoRF: Volumetric Relightable Faces

Pramod Rao[1], Mallikarjun BR[1], Gereon Fox[1], Tim Weyrich[2], Bernd Bickel[3], Hanspeter Pfister[4], Wojciech Matusik[5], Ayush Tewari[5], Christian Theobalt[1], Mohamed Elgharib[1]

[1]Max Planck Institute for Informatics, Saarland Informatics Campus

[2]Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), [3]IST-Austria, [4] Harvard University, [5]MIT CSAIL

max planck institut informatik
UNIVERSITÄT DES SAARLANDES
BMVC 2022

Motivation

- **Portrait viewpoint and illumination editing** is crucial for AR/VR applications.
- Comprehensive **knowledge of geometry and illumination** is critical for obtaining photorealistic results.
- Can we learn a diverse distribution of faces and illuminations to **generalize to unseen subjects** ?
- Can we **accurately model** the **face reflectance** through a **data-driven** approach?

Face Reflectance Fields

We represent **face** as a **volumetric reflectance field** , a volumetric density (σ) and reflectance function (R) pair

$$\sigma(\mathbf{x}): \mathbb{R}^3 \to \mathbb{R}^+ \; ; R(\boldsymbol{\omega}_{\text{in}}, \mathbf{x}, \mathbf{d}): \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R} \text{ s.t.}$$

$$L_{out}(\mathbf{x}, \mathbf{d}) = \int_{\boldsymbol{\omega}_{\text{in}} \in S} R(\boldsymbol{\omega}_{in}, \mathbf{x}, \mathbf{d}) \cdot L_{in}(\boldsymbol{\omega}_{in}) d\omega$$

The **lightstage discretizes** incident light directions S into a finite set I, for $i \in I$ s.t. $S_i \subseteq S$.

$$L_{out}(\mathbf{x}, \mathbf{d}) \approx \sum_{i \in I} R(\boldsymbol{\omega}_i, \mathbf{x}, \mathbf{d}) \cdot L_{in}(\boldsymbol{\omega}_i)$$

OLATs as a light basis allow us to **compose natural lighting conditions.** Under natural illumination conditions:

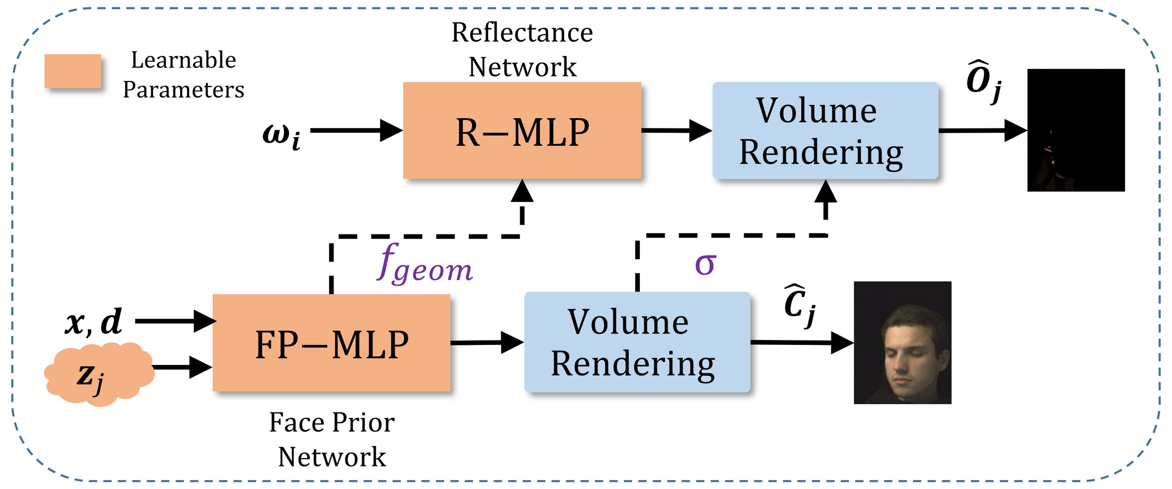
$$\forall i \in I \; \exists \, f_i \text{ s.t. } L_{in}(\omega_i) = f_i \cdot L_{in}(\boldsymbol{\omega}_i)$$

$$L_{out}(\mathbf{x}, \mathbf{d}) \approx \sum_{i \in I} f_i \cdot \int_{t_n}^{t_f} T(t)\, \sigma(\mathbf{x}(t)) R(\boldsymbol{\omega}_i, \mathbf{x}(t), \mathbf{d}) L_{in}(\boldsymbol{\omega}_i) dt$$

$$= \sum_{i \in I} f_i \cdot L(\boldsymbol{\omega}_i, r)$$

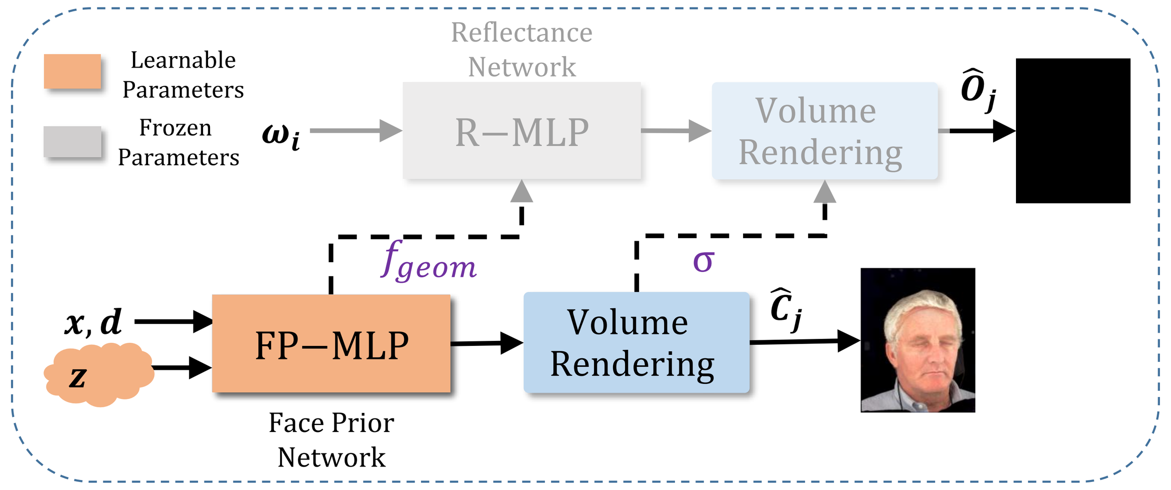
Our Method

Training:



$$\mathcal{L}_{train} = \alpha\, \mathcal{L}_{rgb} + \beta \mathcal{L}_{reg} + \delta \mathcal{L}_h + \delta \mathcal{L}_{\text{OLAT}}$$

Testing:



$$\mathcal{L}_{test} = \alpha\, \mathcal{L}_{rgb} + \beta \mathcal{L}_{reg} + \gamma \mathcal{L}_h$$

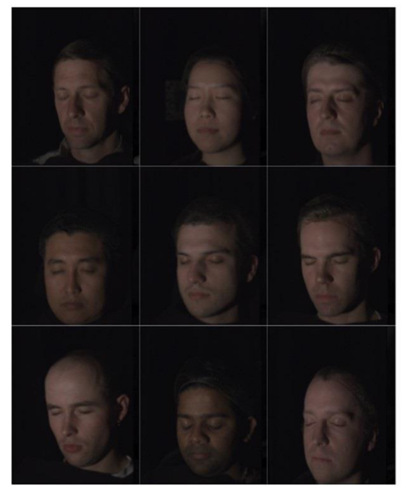
Losses:

$$\mathcal{L}_{rgb} := \sum_{j \in \mathbb{J}} \left(\left\| \hat{\mathbf{C}}_j - \mathbf{C}_j \right\|_2^2 \right) \qquad \mathcal{L}_{reg} := \sum_{j \in \mathbb{J}} \left(\left\| \mathbf{z}_j^{id} \right\|_2^2 + \left\| \mathbf{z}_j^{env} \right\|_2^2 \right)$$

$$\mathcal{L}_h := \sum_{r,k} - \log\left(\mathbb{P}(w_{r,k}) \right) \qquad \mathcal{L}_{\text{OLAT}} := \sum_{j \in \mathbb{J}} \left(\left\| \frac{\hat{\mathbf{O}}_{j,i} - \mathbf{O}_{j,i}}{S(\hat{\mathbf{O}}_{j,i}) + \epsilon} \right\|_2^2 \right)$$

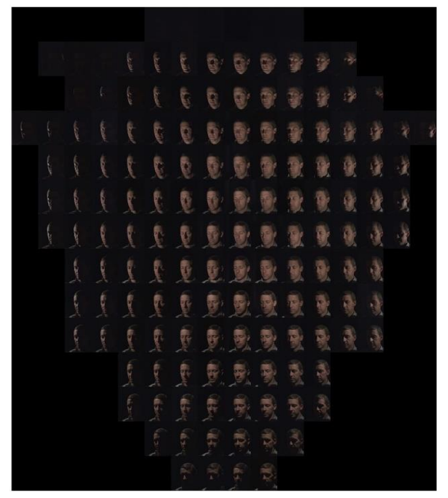
$$where, \mathbb{P}(w_{r,k}) \propto e^{-|w_{r,k}|} + e^{-|1 - w_{r,k}|}$$

Dataset



352 Identities

16 Camera Poses



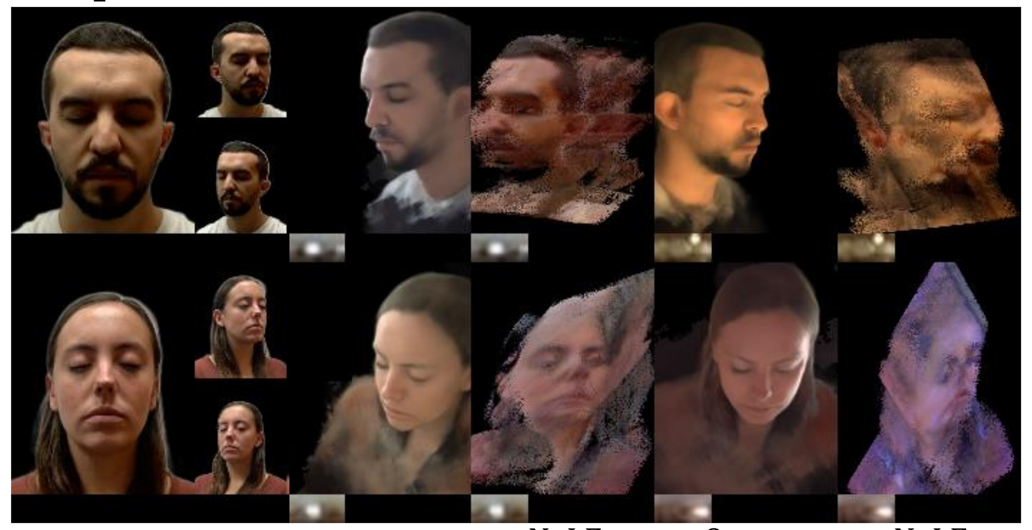
150 Light Sources

Results

Method	Landmark Loss
PhotoApp	2060.25
Ours	1021.62

Input ⟵ NV + RL ⟶

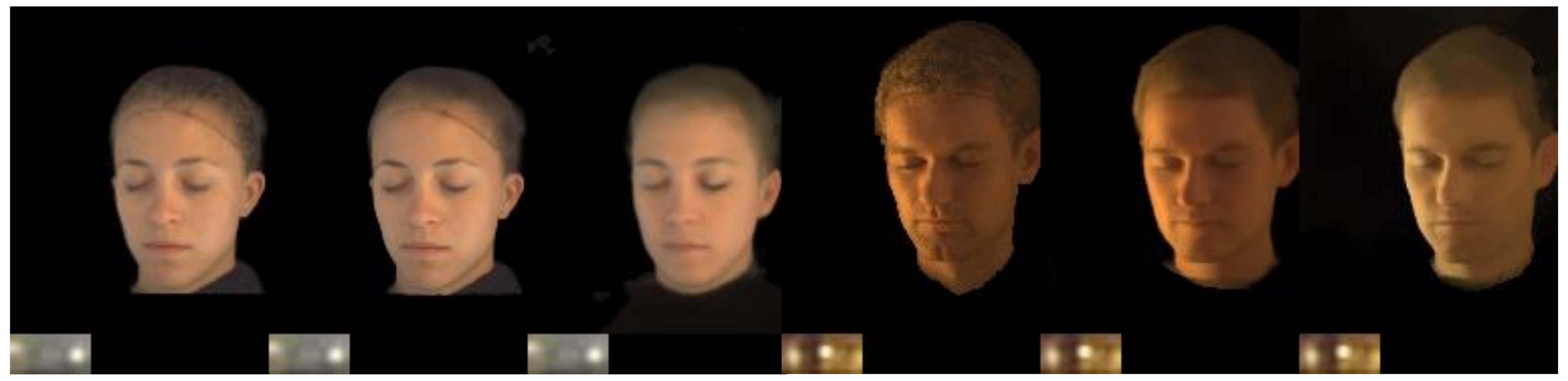
Method	SSIM	PSNR
NeLF	19.72	0.75
Ours	22.80	0.76



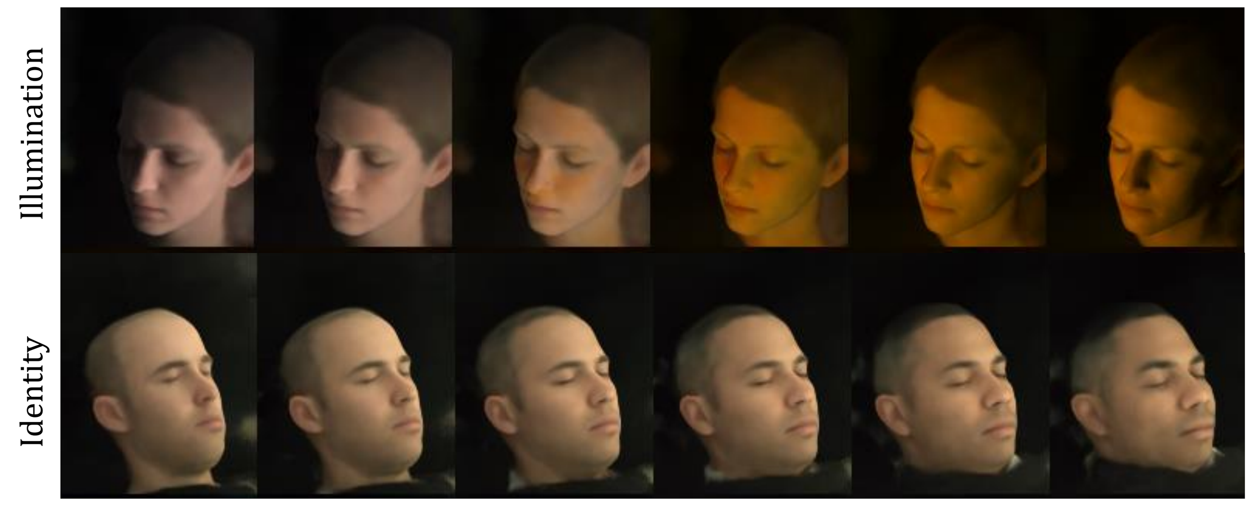
Ours NeLF Ours NeLF

Ablations:

Inputs OLAT Predictions



GT Fit+Finetune Fit Only GT With RN w/o RN



Latent space interpolations

https://vcai.mpi-inf.mpg.de/projects/VoRF/