3D Vision and
Robotics Lab

[ECCV 2022] NeuMan: Neural Human Radiance Field from a Single Video

Wei Jiang†[1,2], Kwang Moo Yi[1,2], Golnoosh Samei[1],
Oncel Tuzel[1], and Anurag Ranjan[1]

[1] Apple
[2] The University of British Columbia
{jw221, kmyi}@cs.ubc.ca, {golnoosh, otuzel, anuragr}@apple.com

Lab Seminar

Gyeongsu Cho @UNIST

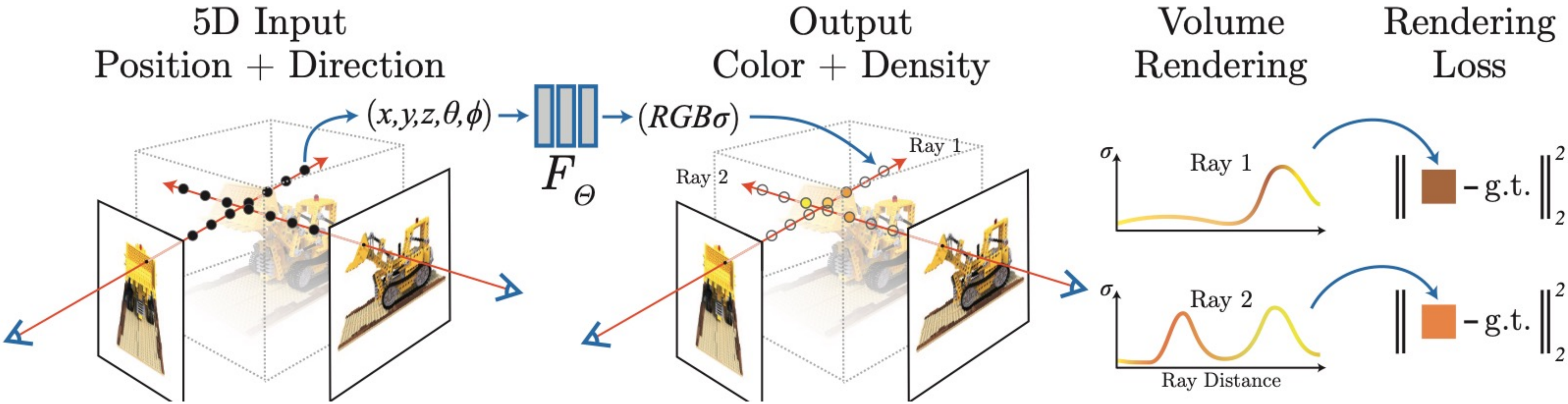
2023.05.04. (Thu)

Contents

Introduction

Introduction

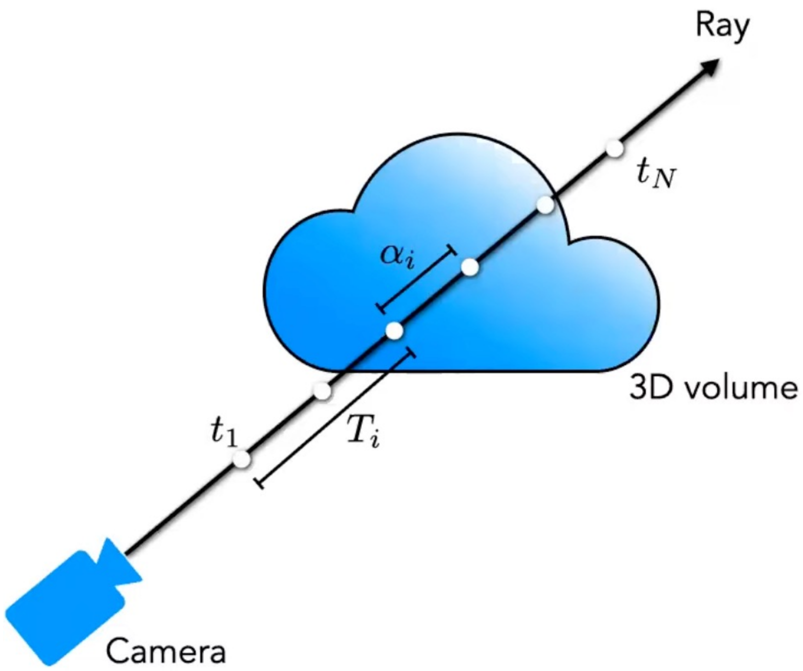
NeRF [1]



[1] Ben Mildenhall, et al, Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV, 2020

Introduction

Volume rendering in NeRF [1]



[1] Ben Mildenhall, et al, Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV, 2020

Introduction

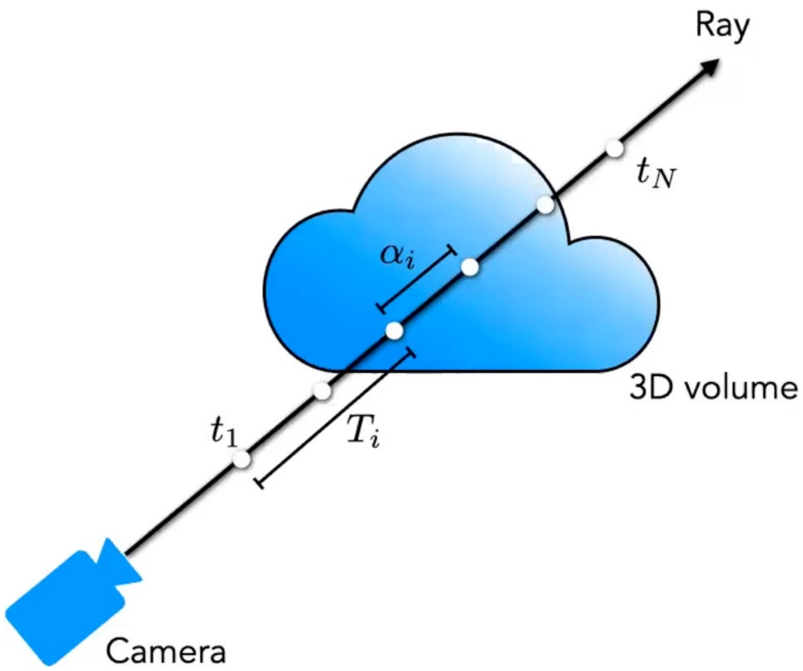
Volume rendering in NeRF [1]

Rendering model for ray r(t) = o + td:

$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i$$ [2]

weights

colors



[1] Ben Mildenhall, et al, Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV, 2020
[2] Max, N.: Optical models for direct volume rendering. In IEEE Transactions on Visualization and Computer Graphics, 1995

Introduction

Volume rendering in NeRF [1]

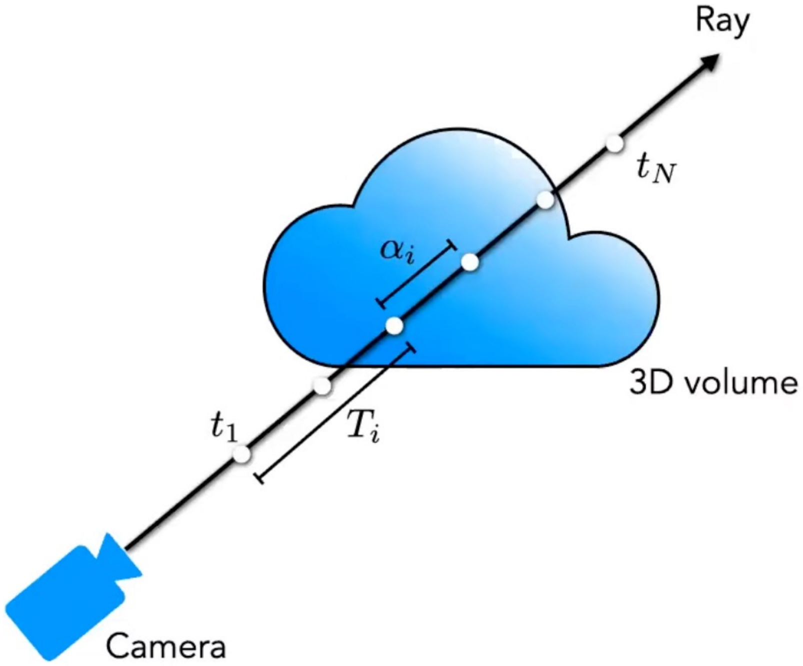
Rendering model for ray r(t) = o + td:

$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i \quad [2]$$

weights: $T_i \alpha_i$; colors: c_i

How much light is blocked earlier along ray:

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$



[1] Ben Mildenhall, et al, Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV, 2020
[2] Max, N.: Optical models for direct volume rendering. In IEEE Transactions on Visualization and Computer Graphics, 1995

Introduction

Volume rendering in NeRF [1]

Rendering model for ray r(t) = o + td:

$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i$$ [2]

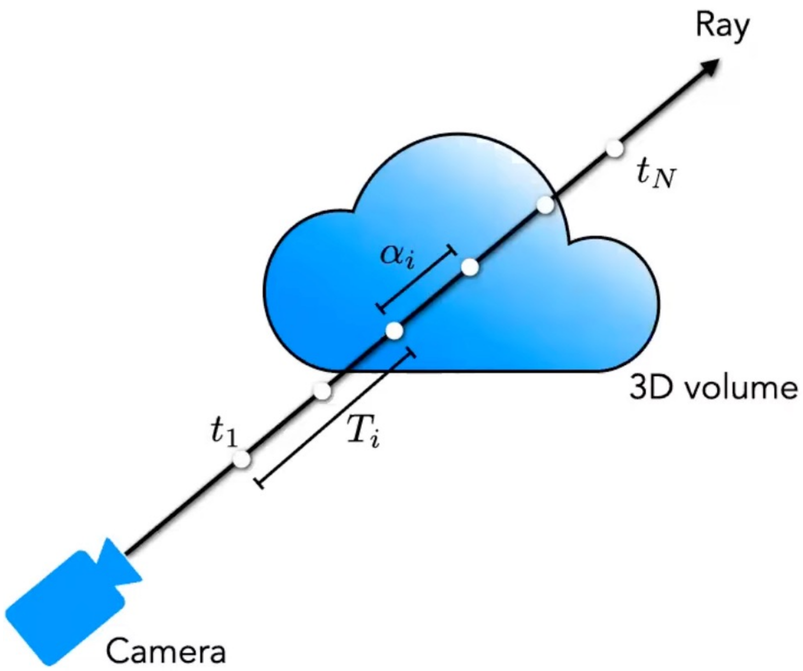
weights ($T_i \alpha_i$), colors (c_i)

How much light is blocked earlier along ray:

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$

How much light is contributed by ray segment i:

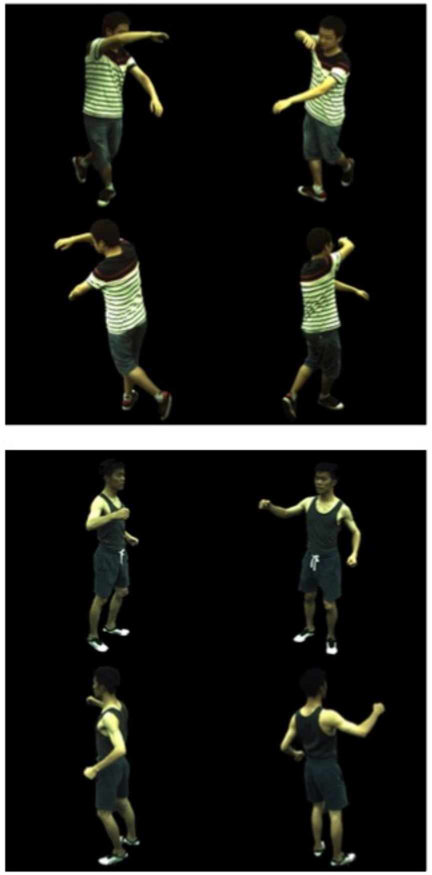
$$\alpha_i = 1 - e^{-\sigma_i \delta t_i}$$



[1] Ben Mildenhall, et al, Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV, 2020
[2] Max, N.: Optical models for direct volume rendering. In IEEE Transactions on Visualization and Computer Graphics, 1995

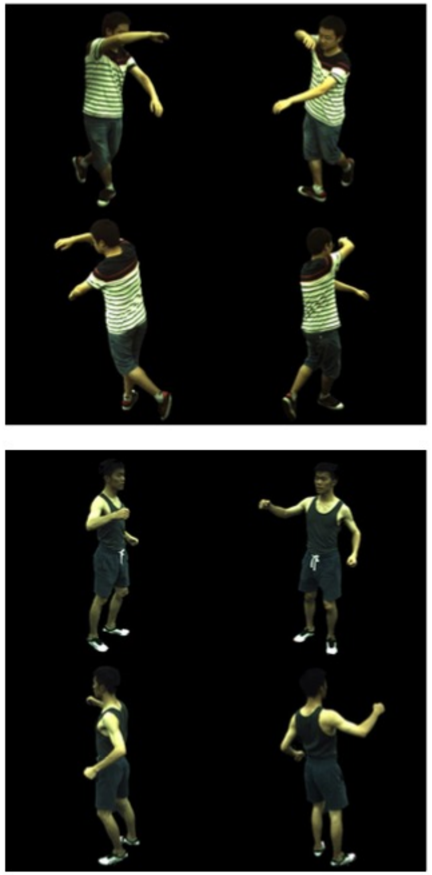
Introduction

[CVPR 2021] Neural Body



Introduction

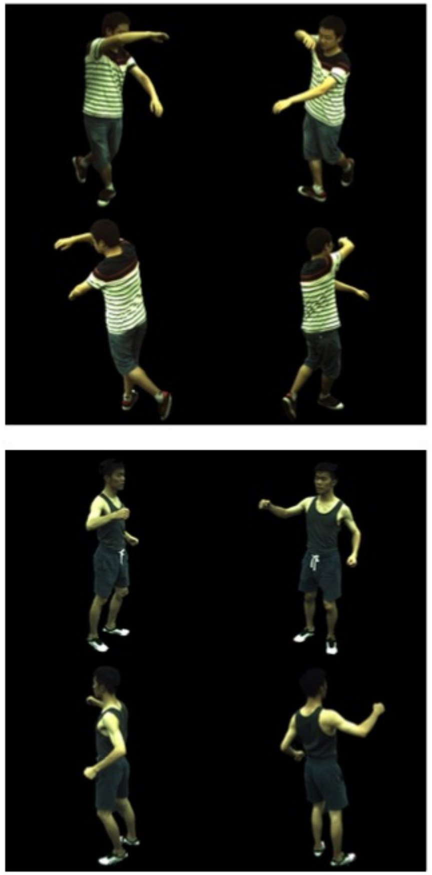
[CVPR 2021] Neural Body



[arXiv 2020] Vid2Actor

Introduction

[CVPR 2021] Neural Body

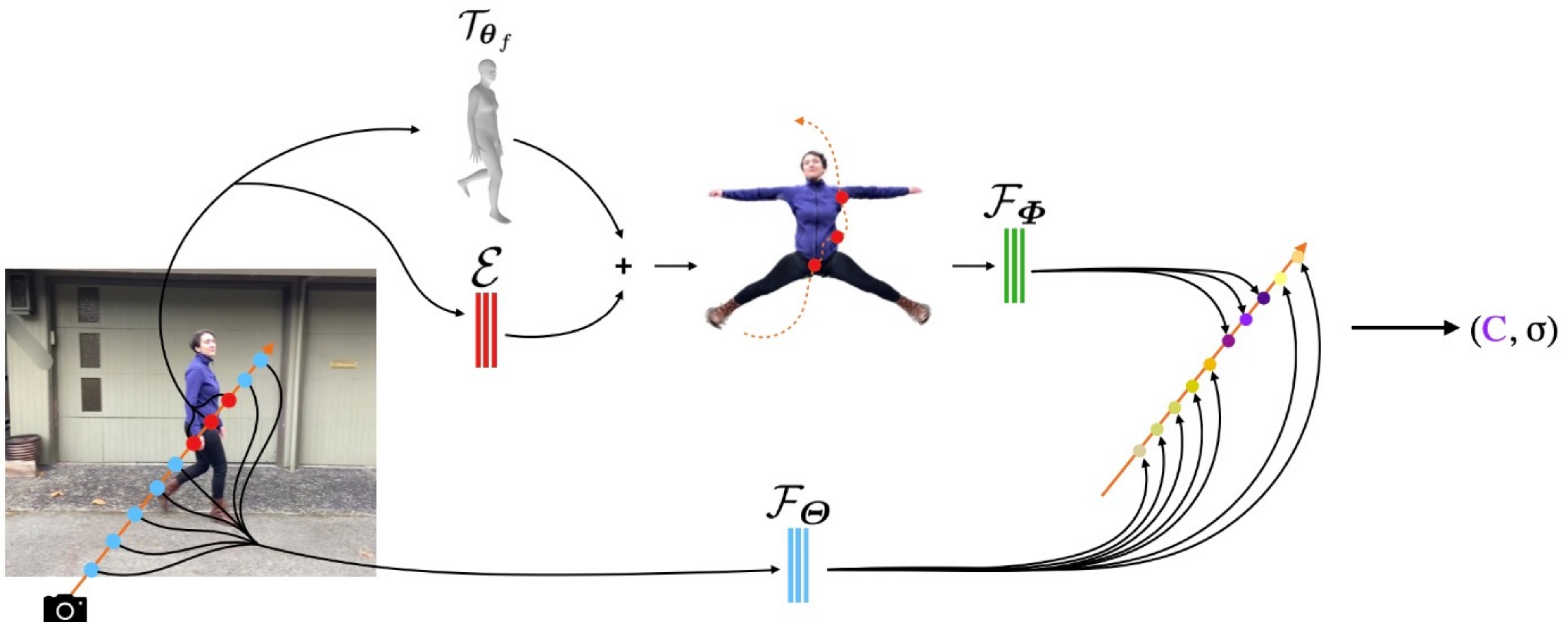


[arXiv 2020] Vid2Actor

[CVPR 2022] HumanNeRF

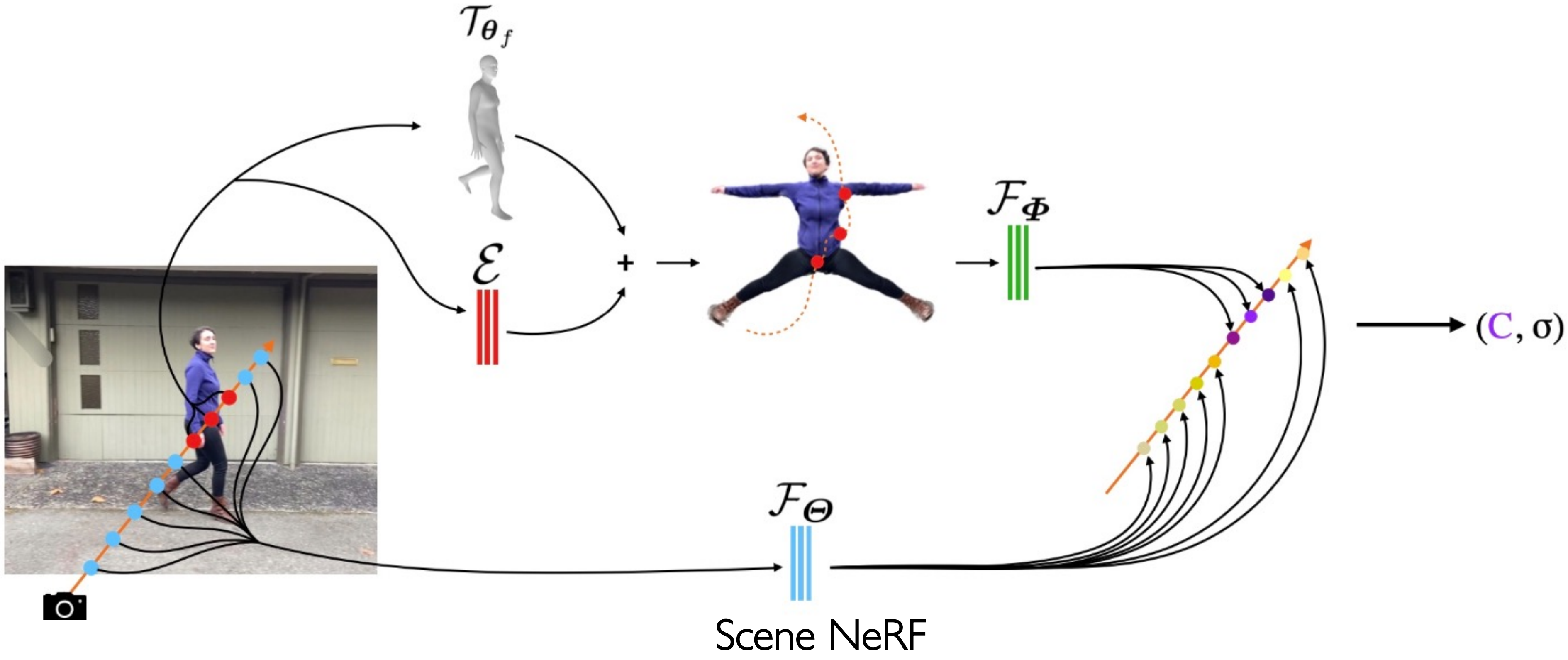
Method

Method



Rendering pipeline

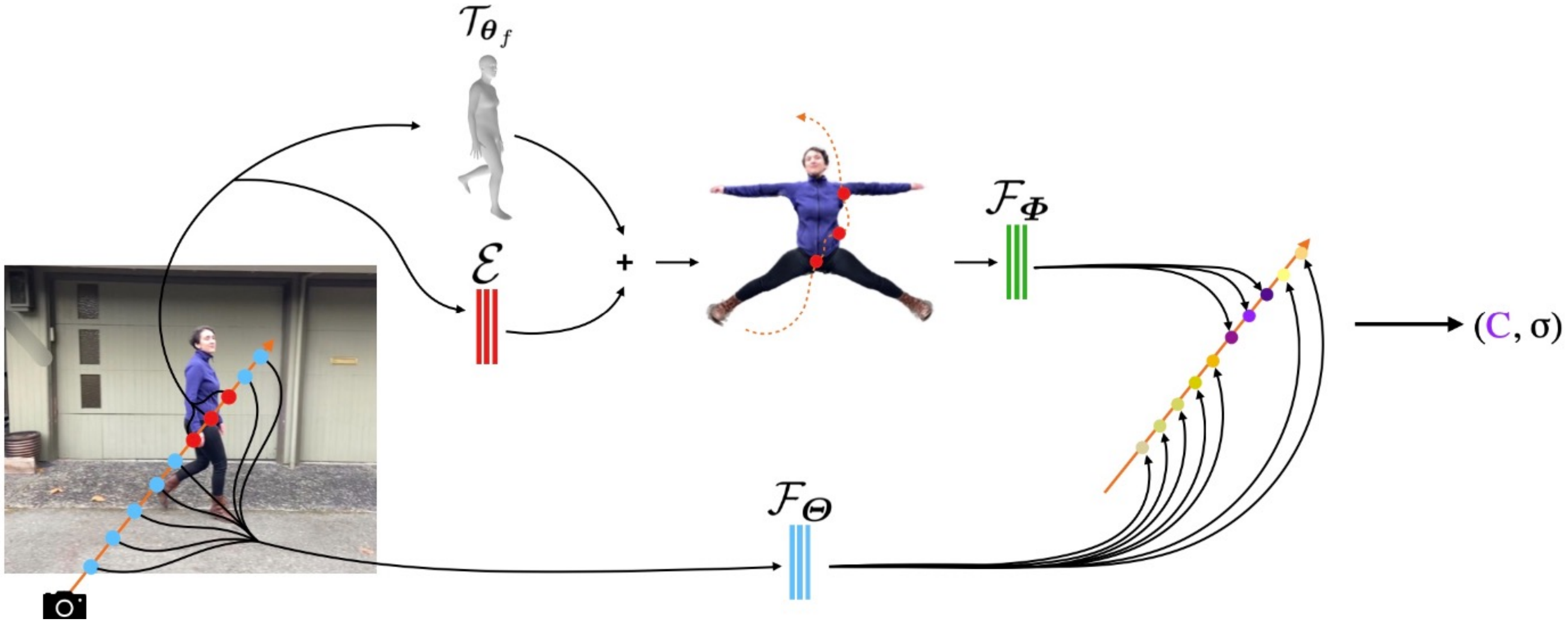
Method



$\mathcal{T}_{\boldsymbol{\theta}_f}$
$\mathcal{E}$
+
$\mathcal{F}_{\boldsymbol{\Phi}}$
$(\mathbf{C}, \sigma)$
$\mathcal{F}_{\boldsymbol{\Theta}}$
Scene NeRF

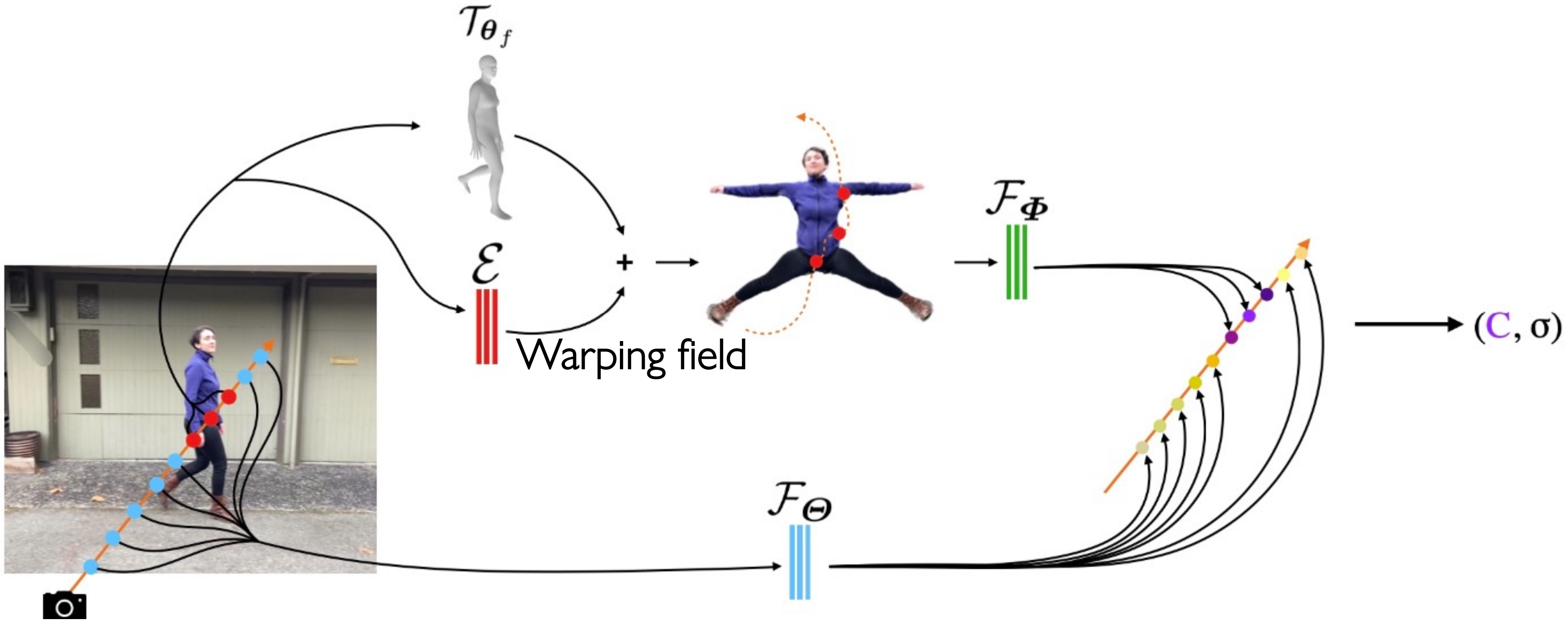
Method

Human Branch



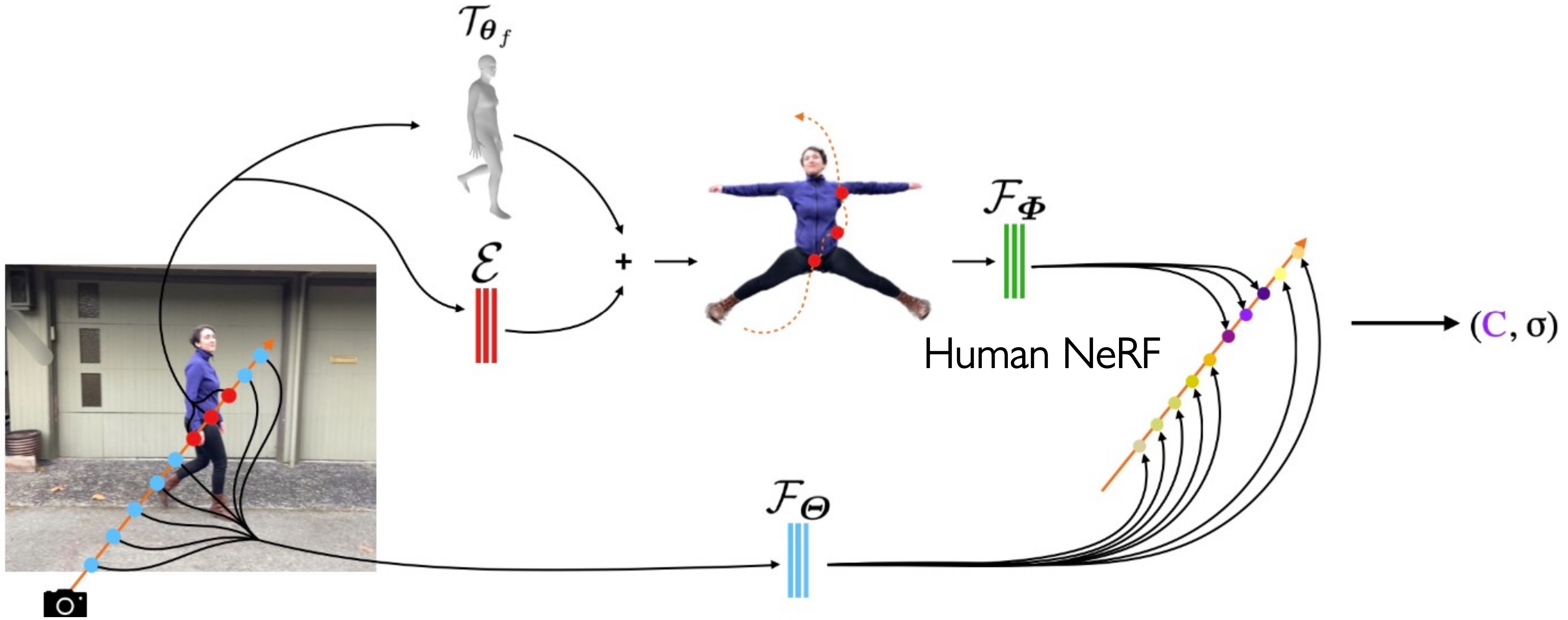
Method

Human Branch



Method

Human Branch



Method

Scene NeRF

Method

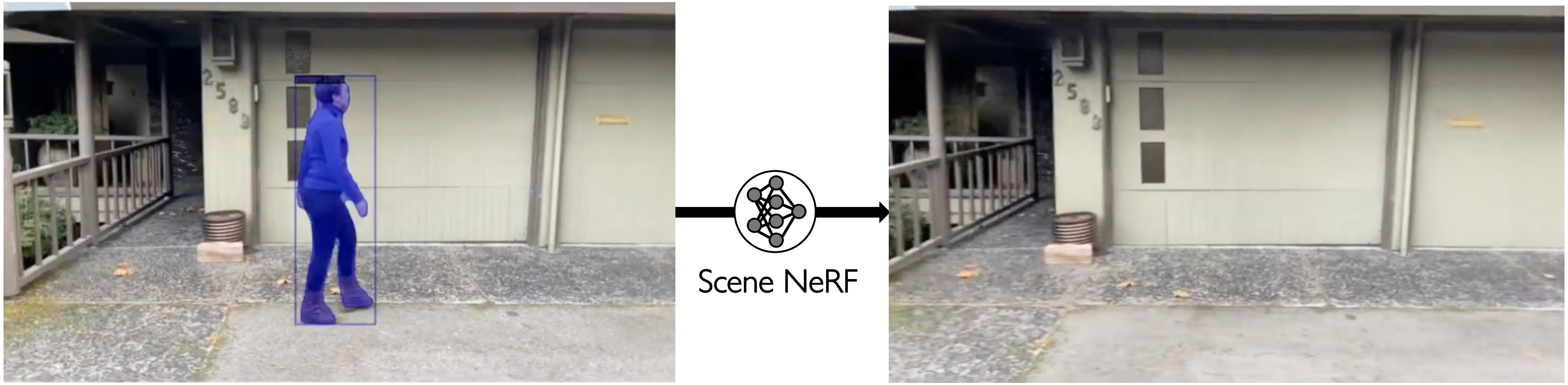
Scene NeRF

Training the scene NeRF model by reconstructing the background pixels detected by Mask-RCNN

Method

Scene NeRF

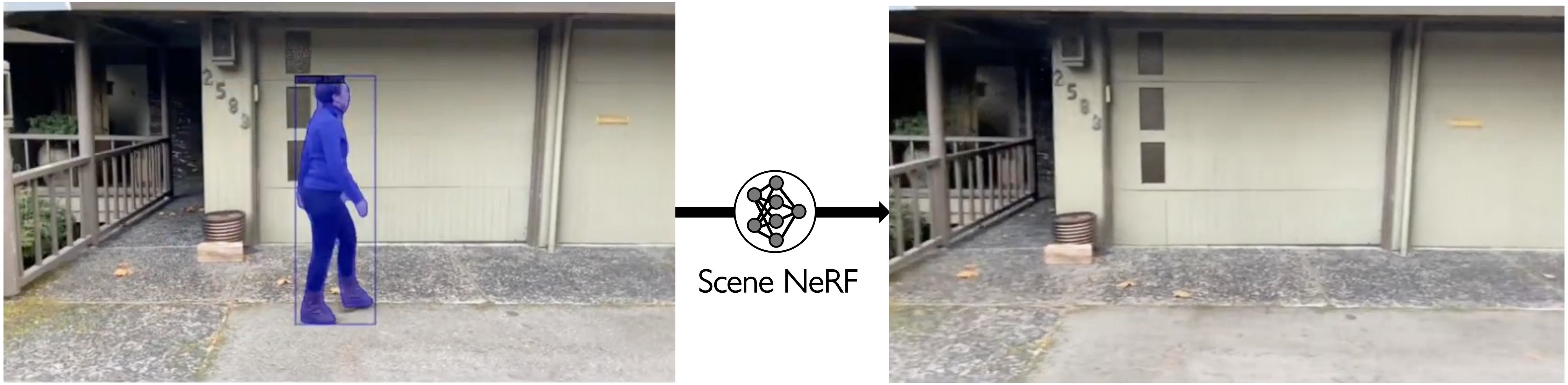
Training the scene NeRF model by reconstructing the background pixels detected by Mask-RCNN



Method

Scene NeRF

Training the scene NeRF model by reconstructing the background pixels detected by Mask-RCNN



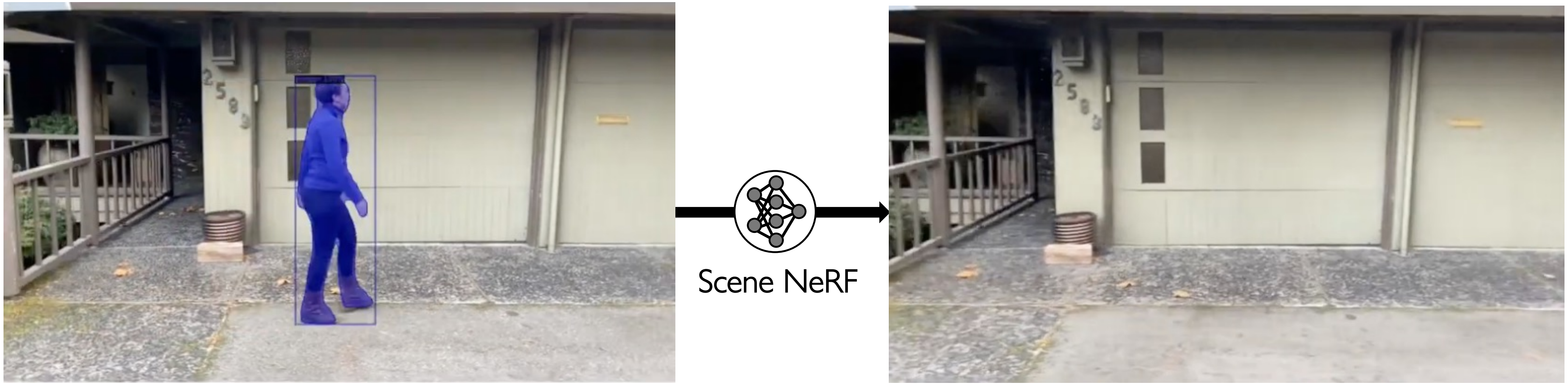
1. RGB Reconstruction Loss

$$\mathcal{L}_{s,rgb}(\mathbf{r}) = (1 - \mathcal{M}(\mathbf{r})) \left\| \mathbf{C}_s(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| ,$$

Method

Scene NeRF

Training the scene NeRF model by reconstructing the background pixels detected by Mask-RCNN



1. RGB Reconstruction Loss

$$\mathcal{L}_{s,rgb}(\mathbf{r}) = (1 - \mathcal{M}(\mathbf{r})) \left\| \mathbf{C}_s(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| \ ,$$

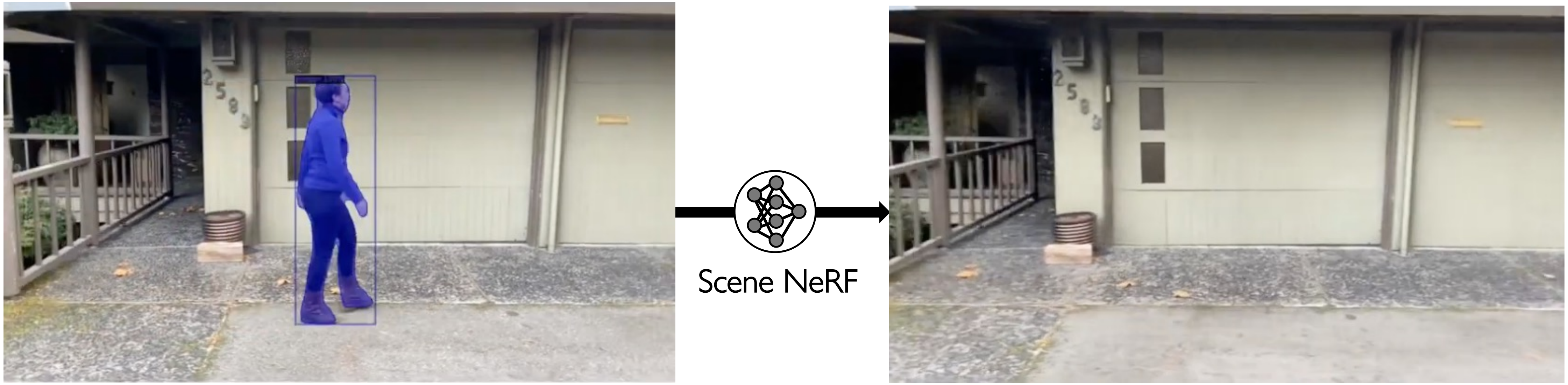
2. Empty Loss

$$\mathcal{L}_{s,empty}(\mathbf{r}) = \int_{t_n}^{\alpha \hat{z}_{\mathbf{r}}} \sigma_s \left(\mathbf{r}(t) \right) dt \ ,$$

Method

Scene NeRF

Training the scene NeRF model by reconstructing the background pixels detected by Mask-RCNN



1. RGB Reconstruction Loss

$$\mathcal{L}_{s,rgb}(\mathbf{r}) = (1 - \mathcal{M}(\mathbf{r})) \left\| \mathbf{C}_s(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| \ ,$$

2. Empty Loss

$$\mathcal{L}_{s,empty}(\mathbf{r}) = \int_{t_n}^{\alpha \hat{z}_{\mathbf{r}}} \sigma_s\left(\mathbf{r}(t)\right) dt \ ,$$

Total Scene Loss

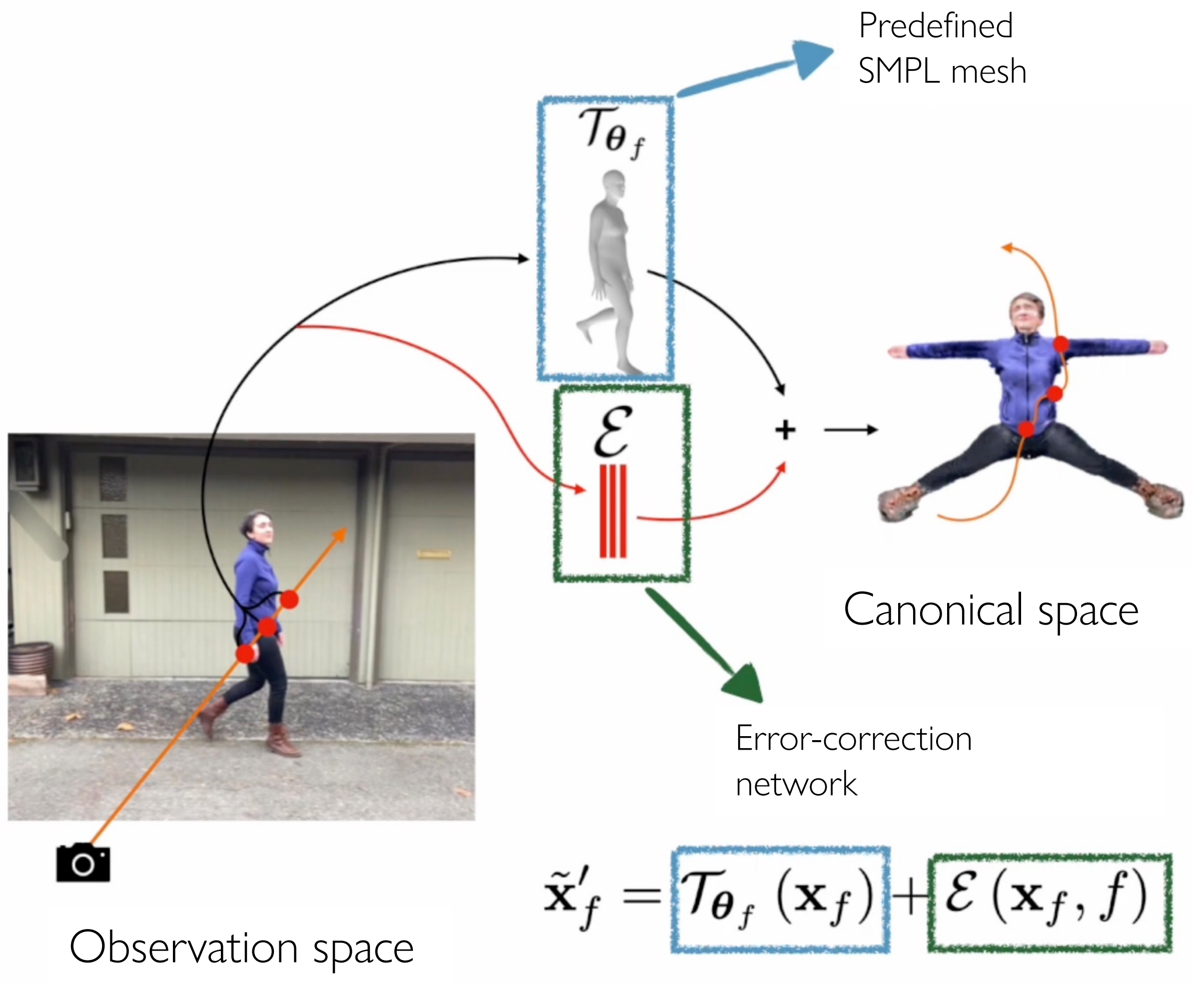
$$\mathcal{L}_s = \mathcal{L}_{s,rgb}(\mathbf{r}) + \lambda_{empty} \mathcal{L}_{s,empty}(\mathbf{r}) \ ,$$

Method

Human Branch – Warping field

Method

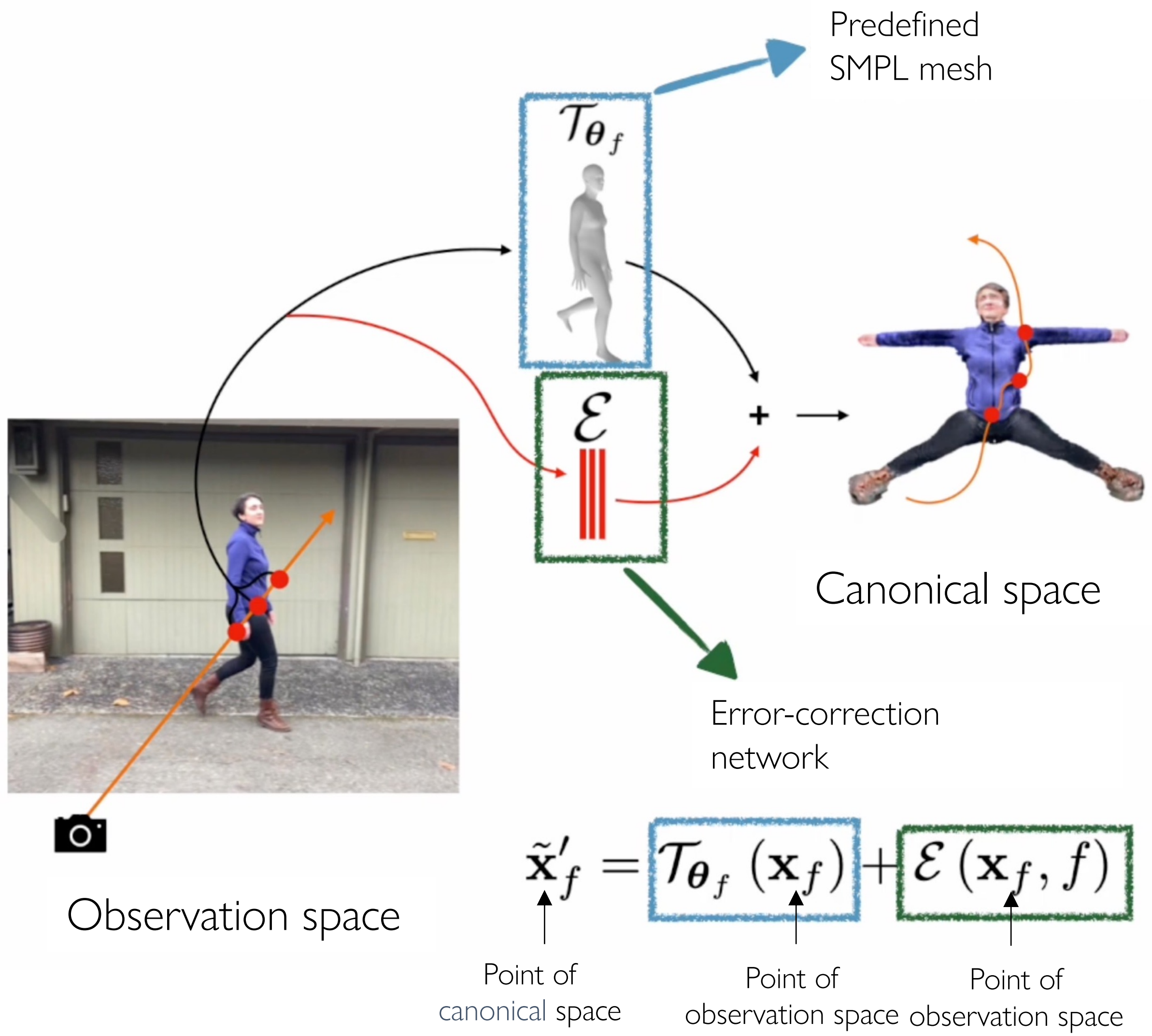
Human Branch – Warping field



$$\tilde{\mathbf{x}}'_f = \mathcal{T}_{\boldsymbol{\theta}_f}(\mathbf{x}_f) + \mathcal{E}(\mathbf{x}_f, f)$$

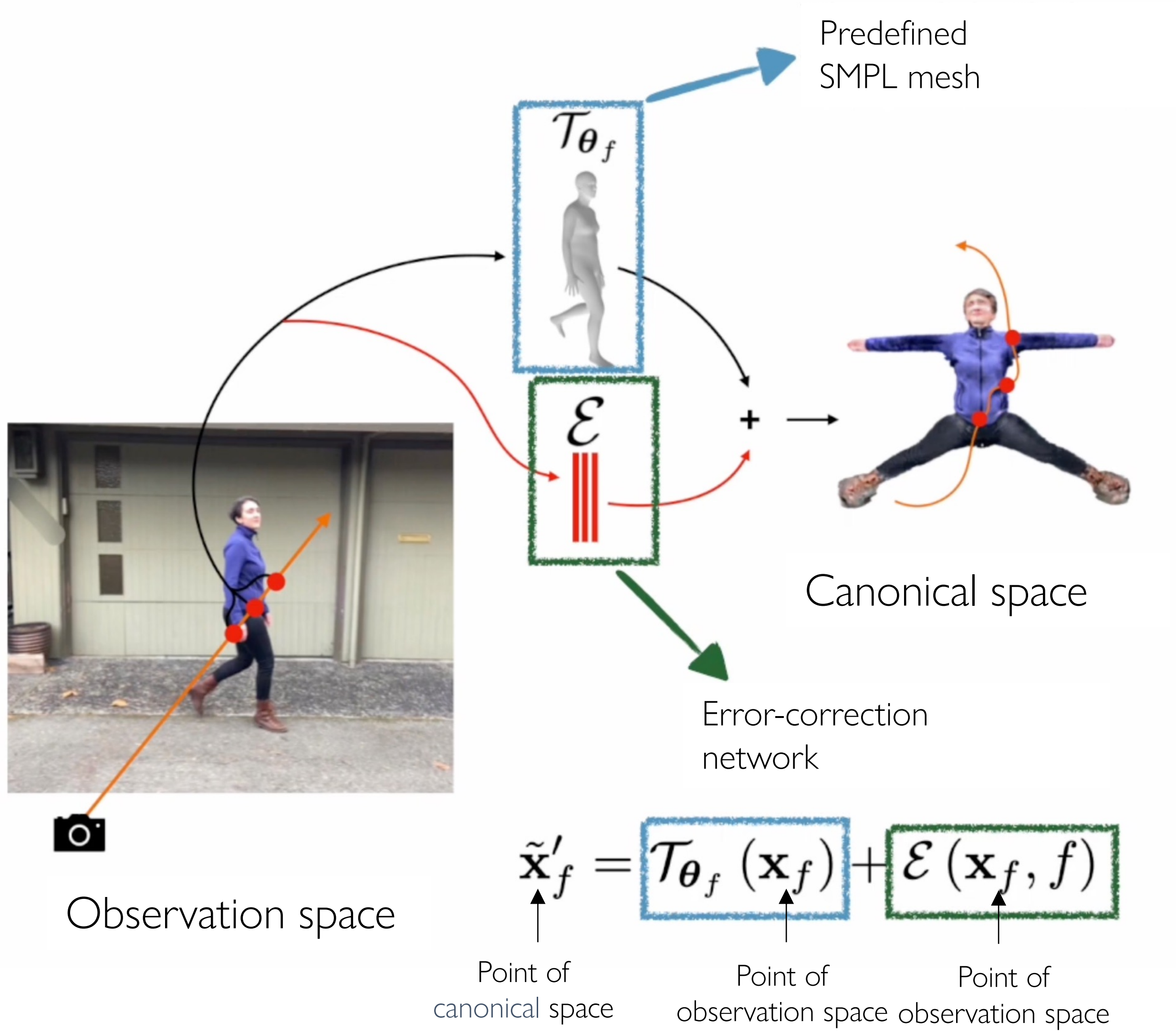
Method

Human Branch – Warping field



Method

Human Branch – Warping field

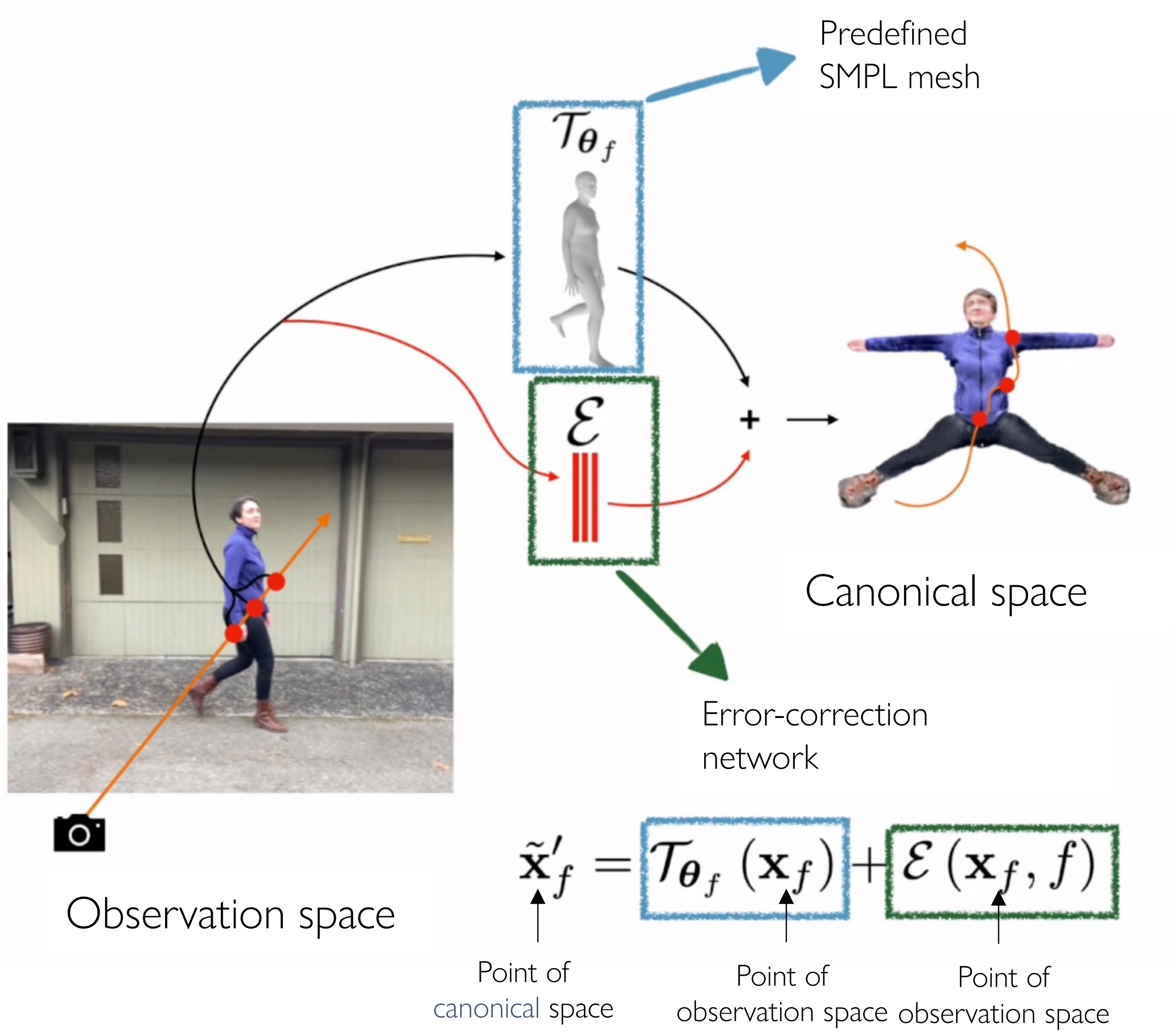


Position

$$\tilde{\mathbf{x}}'_f = \mathcal{T}_{\boldsymbol{\theta}_f}(\mathbf{x}_f) + \mathcal{E}(\mathbf{x}_f, f) \ .$$

Method

Human Branch – Warping field



Position

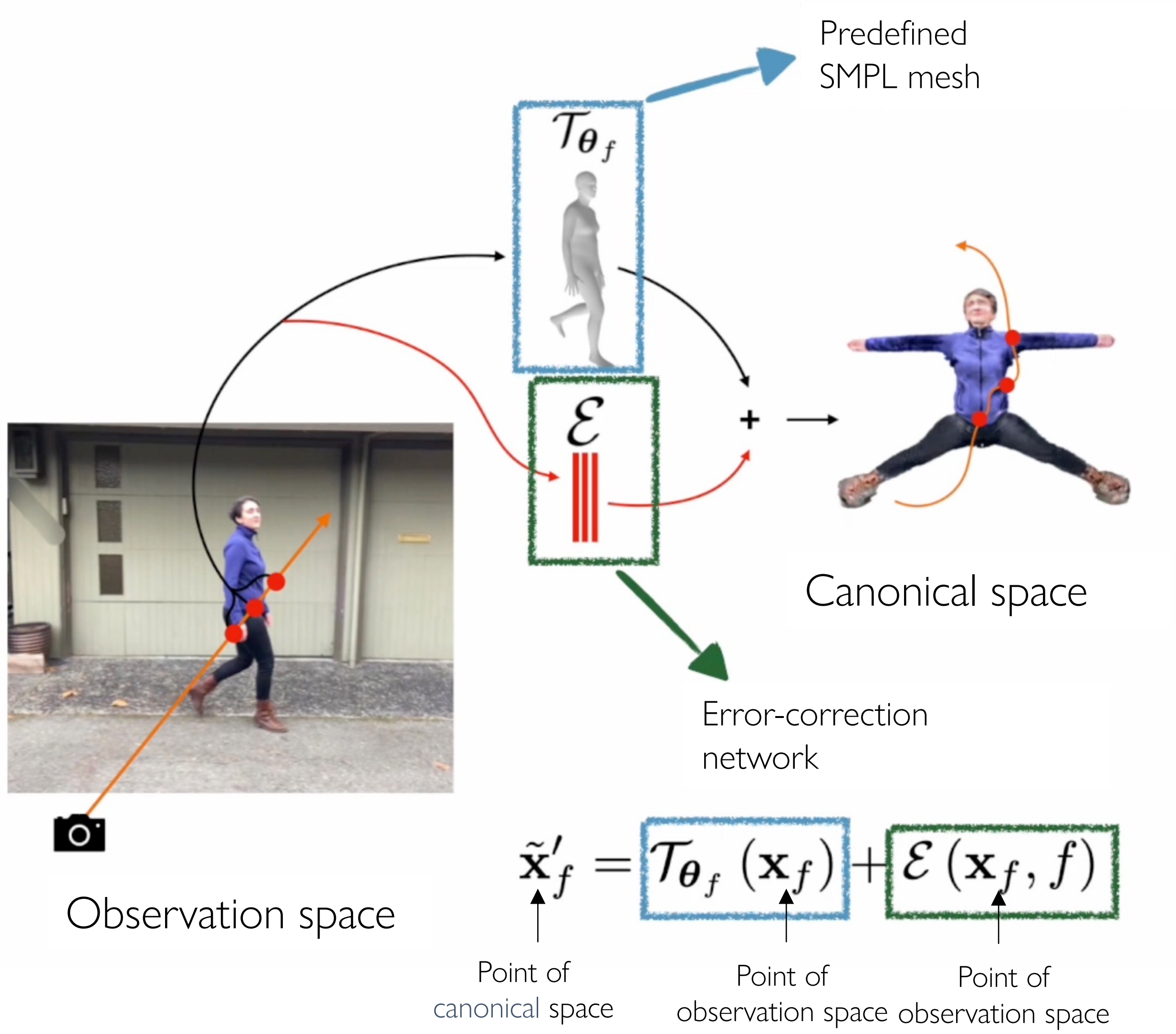
$$\tilde{\mathbf{x}}'_f = \mathcal{T}_{\boldsymbol{\theta}_f}(\mathbf{x}_f) + \mathcal{E}(\mathbf{x}_f, f) \ .$$

Direction

$$\mathbf{d}(t_i)'_f = \frac{\tilde{\mathbf{x}}'_f(t_i) - \tilde{\mathbf{x}}'_f(t_{i-1})}{\left\| \tilde{\mathbf{x}}'_f(t_i) - \tilde{\mathbf{x}}'_f(t_{i-1}) \right\|} \ ,$$

Method

Human Branch – Warping field



Position

$$\tilde{\mathbf{x}}'_f = \mathcal{T}_{\boldsymbol{\theta}_f}(\mathbf{x}_f) + \mathcal{E}(\mathbf{x}_f, f) \; .$$

Direction

$$\mathbf{d}(t_i)'_f = \frac{\tilde{\mathbf{x}}'_f(t_i) - \tilde{\mathbf{x}}'_f(t_{i-1})}{\left\|\tilde{\mathbf{x}}'_f(t_i) - \tilde{\mathbf{x}}'_f(t_{i-1})\right\|} \; ,$$

Human NeRF

$$\mathbf{c}_h, \sigma_h = \mathcal{F}_{\boldsymbol{\Phi}}(\tilde{\mathbf{x}}'_f, \mathbf{d}'_f) \; ,$$

Method

Human Branch – Human NeRF

Method

Human Branch – Human NeRF

1. RGB Reconstruction Loss

$$\mathcal{L}_{h,rgb}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| \mathbf{C}_h(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| ,$$

Method

Human Branch – Human NeRF

1. RGB Reconstruction Loss

$$\mathcal{L}_{h,rgb}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| \mathbf{C}_h(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| \ ,$$

2. Mask Loss

$$\mathcal{L}_{mask}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| 1 - \alpha_h(\mathbf{r}) \right\| \ ,$$

Method

Human Branch – Human NeRF

1. RGB Reconstruction Loss

$$\mathcal{L}_{h,rgb}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| \mathbf{C}_h(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| \ ,$$

2. Mask Loss

$$\mathcal{L}_{mask}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| 1 - \alpha_h(\mathbf{r}) \right\| \ ,$$

3. SMPL Loss

$$\mathcal{L}_{smpl}(\hat{\mathbf{x}}'_f, \sigma_h) = \begin{cases} \|1 - \sigma_h\| \,, & \text{if } \hat{\mathbf{x}}'_f \text{ inside SMPL mesh} \\ |\sigma_h|, & \text{otherwise} \end{cases} \ ,$$

Method

Human Branch – Human NeRF

1. RGB Reconstruction Loss

$$\mathcal{L}_{h,rgb}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| \mathbf{C}_h(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| \ ,$$

2. Mask Loss

$$\mathcal{L}_{mask}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| 1 - \alpha_h(\mathbf{r}) \right\| \ ,$$

3. SMPL Loss

$$\mathcal{L}_{smpl}(\hat{\mathbf{x}}'_f, \sigma_h) = \begin{cases} \|1 - \sigma_h\| , & \text{if } \hat{\mathbf{x}}'_f \text{ inside SMPL mesh} \\ |\sigma_h|, & \text{otherwise} \end{cases} \ ,$$

4. Hard Loss

$$\mathcal{L}_{hard} = -\log(e^{-|w|} + e^{-|1-w|})$$

Method

Human Branch – Human NeRF

1. RGB Reconstruction Loss

$$\mathcal{L}_{h,rgb}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| \mathbf{C}_h(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| \ ,$$

2. Mask Loss

$$\mathcal{L}_{mask}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| 1 - \alpha_h(\mathbf{r}) \right\| \ ,$$

3. SMPL Loss

$$\mathcal{L}_{smpl}(\hat{\mathbf{x}}'_f, \sigma_h) = \begin{cases} \|1 - \sigma_h\|, & \text{if } \hat{\mathbf{x}}'_f \text{ inside SMPL mesh} \\ |\sigma_h|, & \text{otherwise} \end{cases} \ ,$$

4. Hard Loss

$$\mathcal{L}_{hard} = -\log(e^{-|w|} + e^{-|1-w|})$$

[CVPR 2022] LoLNeRF

Method

Human Branch – Human NeRF

1. RGB Reconstruction Loss

$$\mathcal{L}_{h,rgb}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| \mathbf{C}_h(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| \ ,$$

2. Mask Loss

$$\mathcal{L}_{mask}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| 1 - \alpha_h(\mathbf{r}) \right\| \ ,$$

3. SMPL Loss

$$\mathcal{L}_{smpl}(\hat{\mathbf{x}}'_f, \sigma_h) = \begin{cases} \|1 - \sigma_h\|, & \text{if } \hat{\mathbf{x}}'_f \text{ inside SMPL mesh} \\ |\sigma_h|, & \text{otherwise} \end{cases} \ ,$$

4. Hard Loss

$$\mathcal{L}_{hard} = -\log(e^{-|w|} + e^{-|1-w|})$$

5. Edge Loss

$$\mathcal{L}_{edge} = -\log(e^{-|\alpha_c|} + e^{-|1-\alpha_c|})$$

Method

Human Branch – Human NeRF

1. RGB Reconstruction Loss

$$\mathcal{L}_{h,rgb}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| \mathbf{C}_h(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r}) \right\| ,$$

2. Mask Loss

$$\mathcal{L}_{mask}(\mathbf{r}) = \mathcal{M}(\mathbf{r}) \left\| 1 - \alpha_h(\mathbf{r}) \right\| ,$$

3. SMPL Loss

$$\mathcal{L}_{smpl}(\hat{\mathbf{x}}'_f, \sigma_h) = \begin{cases} \|1 - \sigma_h\|, & \text{if } \hat{\mathbf{x}}'_f \text{ inside SMPL mesh} \\ |\sigma_h|, & \text{otherwise} \end{cases} ,$$

4. Hard Loss

$$\mathcal{L}_{hard} = -\log(e^{-|w|} + e^{-|1-w|})$$

5. Edge Loss

$$\mathcal{L}_{edge} = -\log(e^{-|\alpha_c|} + e^{-|1-\alpha_c|})$$

Total Loss

$$\mathcal{L} = \mathcal{L}_{h,rgb} + \lambda_{lpips}\mathcal{L}_{lpips} + \lambda_{mask}\mathcal{L}_{mask} + \lambda_{smpl}\mathcal{L}_{smpl} + \lambda_{hard}\mathcal{L}_{hard} + \lambda_{edge}\mathcal{L}_{edge}.$$

Experiments

Experiments

- Dataset
  - Neuman Dataset

- Evaluation metrics
  - PSNR
  - SSIM
  - LPIPS

Experiments

- Dataset
  - Neuman Dataset

Sequence	Total Frames	Train Frames	Validatation Frames	Test Frames
Seattle	41	33	4	4
Citron	37	30	4	3
Parking	42	34	4	4
Bike	104	83	11	10
Jogging	102	82	10	10
Lab	103	82	11	10

- Evaluation metrics
  - PSNR
  - SSIM
  - LPIPS

Experiments

Novel view synthesis comparison

	Seattle			Citron			Parking		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-T	21.84	0.69	0.37	12.33	0.49	0.65	21.98	0.69	0.46
HyperNeRF	16.43	0.43	0.40	16.81	0.41	0.56	16.04	0.38	0.62
Ours	**23.98**	**0.77**	**0.26**	**24.71**	**0.80**	**0.26**	**25.43**	**0.79**	**0.31**

	Bike			Jogging			Lab		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-T	21.16	0.71	0.36	20.63	0.53	0.49	20.52	0.75	0.39
HyperNeRF	17.64	0.42	0.43	18.52	0.39	0.52	16.75	0.51	0.23
Ours	**25.52**	**0.82**	**0.23**	**22.68**	**0.67**	**0.32**	**24.93**	**0.85**	**0.21**

NeRF-T HyperNeRF Ours Ground Truth

Experiments

Human NeRF Reconstructions

Experiments

Ablation Studies

Conclusion

Conclusion