第8回定期ミーティング

2024/10/22

早稲田大学　基幹理工学研究科
電子物理システム学専攻 史研究室
石黒将太郎・野口颯汰

アウトライン

- 論文紹介
  - DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models
- 進捗状況
  - 研究方針
  - データセット

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

目的

- DDMによって学習された色と深度マップの事前知識によって、再構成された幾何構造の品質と、視点の一貫性が向上する

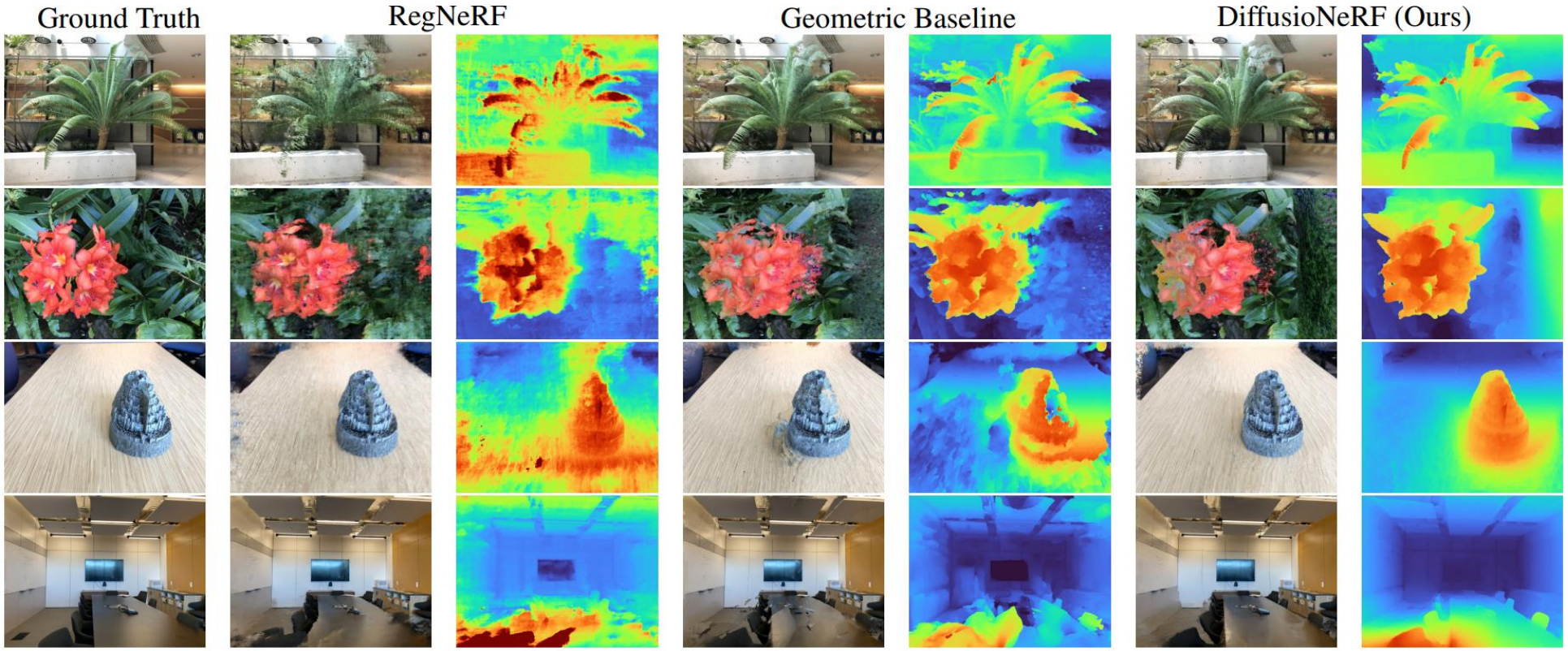

図. モデルのレンダリング結果[1]

[1] Jamie Wynn, Daniyar Turmukhambetov, ” DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models”, CVPR 2023, 2023-11-8

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

全体アーキテクチャ

- トレーニング画像から密度σ・ピクセル色cを予測し、ボリュームレンダリングした結果から誤差逆伝播で重み更新
- Hyperismデータセットを使用し、48×48ピクセルの画像と深度マップをサンプルすることでDDMを訓練

図. Hyperism データセット[2]

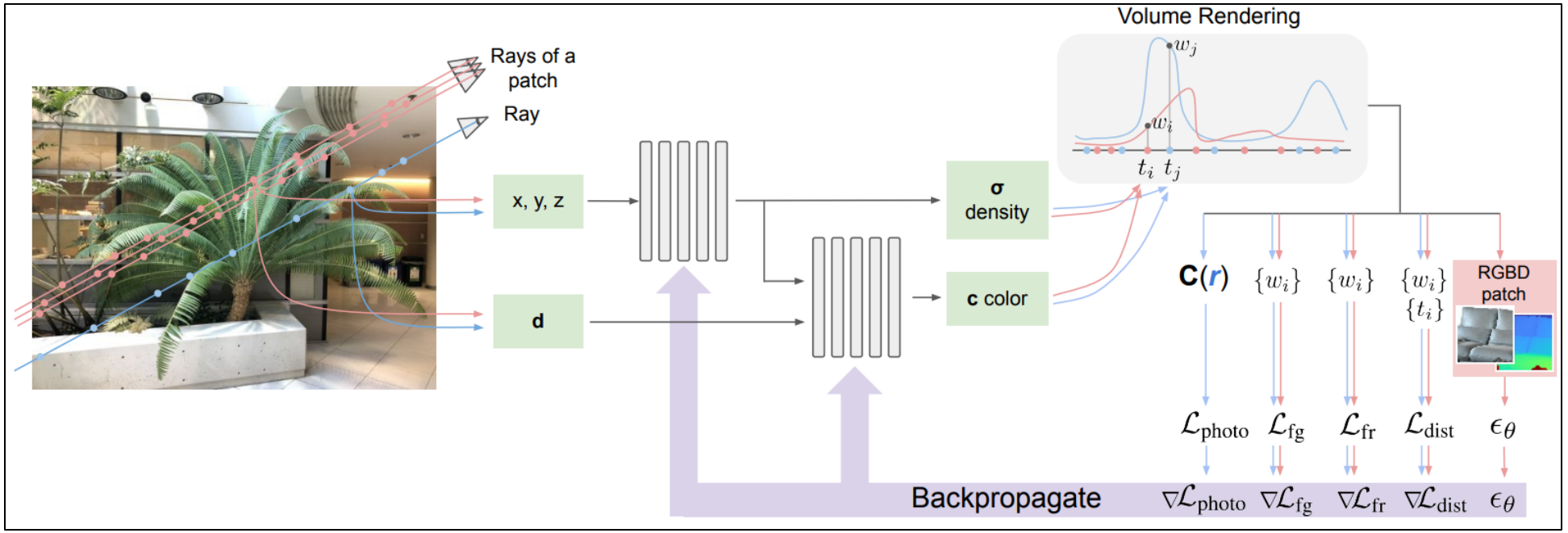

図. 全体アーキテクチャ[1]

[1] Jamie Wynn, Daniyar Turmukhambetov, " DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models", CVPR 2023, 2023-11-8
[2] Mike Roberts AND Jason Ramapuram, "Hypersim: Photorealistic Synthetic Dataset for Holistic Indoor Scene Understanding", ICCV 2021

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

損失関数

- レンダリング

光線$r(t) = o + td$の予測色$C(r)$は光線上の離散的なサンプルt_i：N個を用いて

$$\mathbf{C}(\mathbf{r}) \approx \sum_{i=1}^{N} w_i \mathbf{c}(\mathbf{r}(t_i), \mathbf{d}) + \left(1 - \sum_{i=1}^{N} w_i\right) \mathbf{c}_{\text{bg}}$$

ここで、色の寄与の重みw_iは、サンプルt_iに到達するまでに光線が吸収されない確率と、t_iでの密度に基づく透過率の積
また、光線のどの位置で物体とぶつかるかを表す予測深度$D(r)$は、

$$\mathbf{D}(\mathbf{r}) = \frac{\sum_{i=1}^{N} w_i t_i}{\sum_{i=1}^{N} w_i}.$$

- L_{photo}

L_{photo}は入力画像と予測色のフォトメトリック再構成損失

$$\mathcal{L}_{\text{photo}}(\sigma, \mathbf{c}) = \sum_{i=1}^{\mathcal{I}} ||I_i - \mathbf{C}_i||_2.$$

[1] Jamie Wynn, Daniyar Turmukhambetov, ” DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models”, CVPR 2023, 2023-11-8

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

損失関数

- L_{dist}
  密度フィールドが広がりすぎることによるアーティファクトを防ぐために、σがカメラに近いところで集中するための損失関数

$$\mathcal{L}_{\text{dist}} = \frac{1}{D(\mathbf{r})}\left(\sum_{i,j} w_i w_j \left|\frac{t_i + t_{i+1}}{2} - \frac{t_j + t_{j+1}}{2}\right| + \frac{1}{3}\sum_{i=1}^{N} w_i^2 (t_{i+1} - t_i)\right)$$

- L_{fg}
  光線がジオメトリによって完全に吸収されることを保証するための損失関数

$$\mathcal{L}_{\text{fg}} = \left(1 - \sum_{i=1}^{N} w_i\right)^2 .$$

- L_{fr}
  各トレーニング画像専用の退化解の生成を防ぐ損失関数

$$\mathcal{L}_{\text{fr}} = \sum_i w_i \mathbf{1}(n_i <= 1)$$

DDMノイズ推定器によるスコア関数

$$\nabla\mathcal{L} = \nabla\mathcal{L}_{\text{photo}} + \lambda_{\text{fg}}\nabla\mathcal{L}_{\text{fg}} + \lambda_{\text{fr}}\nabla\mathcal{L}_{\text{fr}} + \lambda_{\text{dist}}\nabla\mathcal{L}_{\text{dist}} - \lambda_{\text{DDM}}\epsilon_\theta$$

[1] Jamie Wynn, Daniyar Turmukhambetov, ” DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models”, CVPR 2023, 2023-11-8

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

学習方法

- アーキテクチャ
  - NeRF：torch-ngpで実装されるInstant NGPを使用
  - Cuda：tiny-cuda-nn

- 最適化方法
  - 12000ステップで最適化
  - 最初の2500ステップのみ$\lambda_{dist} = 0$
- 学習環境
  - Nvidia A100を1台使用
  - シーンごとに30分かかる
- データセット
  - 学習用：hyperism
  - 評価用：LLFF・DTU

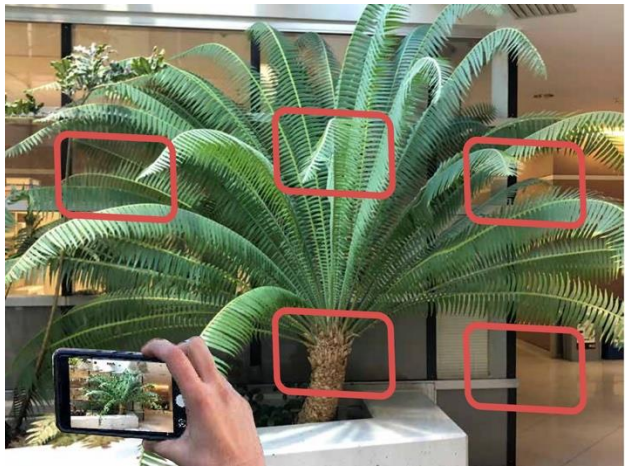

図. LLFF データセット[3]

図. DTU データセット[4]

[1] Jamie Wynn, Daniyar Turmukhambetov, ” DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models”, CVPR 2023, 2023-11-8
[3] Ben Mildenhall and Pratul P. Srinivasan, Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines, TOG 2019
[4] Jensen, Rasmus and Dahl, Anders and Vogiatzis, Large scale multi-view stereopsis evaluation, IEEE 2014

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

評価

- 評価数値上はDDMによるスコア関数なしverが最も高いが、DiffusionNeRFよりアーティファクトが多い

	Method	Setting	PSNR ↑			SSIM ↑			LPIPS-VGG ↓			Average ↓		
			3-view	6-view	9-view	3-view	6-view	9-view	3-view	6-view	9-view	3-view	6-view	9-view
LLFF	mip-NeRF [1]	Optimized per Scene	14.62	20.87	24.26	0.351	0.692	0.805	0.495	0.255	0.172	0.246	0.114	0.073
	DietNeRF [11]	Optimized per Scene	14.94	21.75	24.28	0.370	0.717	0.801	0.496	0.248	0.183	0.240	0.105	0.073
	PixelNeRF ft [45]	DTU + ft per Scene	16.17	17.03	18.92	0.438	0.473	0.535	0.512	0.477	0.430	0.217	0.196	0.163
	MVSNeRF ft [3]	DTU + ft per Scene	17.88	19.99	20.47	0.584	0.660	0.695	0.327	0.264	0.244	0.157	0.122	0.111
	RegNeRF [21]	Optimized per Scene	19.08	21.10	24.86	0.587	0.760	**0.820**	0.336	**0.206**	**0.161**	0.146	0.086	0.067
	Geometric Baseline	Optimized per Scene	**19.88**	**24.28**	**25.10**	**0.590**	**0.765**	0.802	**0.312**	0.210	0.189	**0.129**	**0.076**	**0.066**
	DiffusioNeRF (Ours)	Optimized per Scene	19.79	23.79	25.02	0.568	0.747	0.785	0.338	0.237	0.212	0.136	0.083	0.071
DTU	mip-NeRF [1]	Optimized per Scene	8.68	16.54	23.58	0.571	0.741	0.879	0.353	0.198	0.092	0.323	0.148	0.056
	DietNeRF [11]	Optimized per Scene	11.85	20.63	23.83	0.633	0.778	0.823	0.314	0.201	0.173	0.243	0.101	0.068
	PixelNeRF ft [45]	DTU + ft per Scene	**18.95**	20.56	21.83	0.710	0.753	0.781	0.269	0.223	0.203	0.125	0.104	0.090
	MVSNeRF ft [3]	DTU + ft per Scene	18.54	20.49	22.22	**0.769**	0.822	0.853	0.197	0.155	0.135	0.113	0.089	0.069
	RegNeRF [21]	Optimized per Scene	18.89	**22.20**	24.93	0.745	**0.841**	**0.884**	**0.190**	**0.117**	**0.089**	**0.112**	**0.071**	**0.047**
	Geometric Baseline	Optimized per Scene	13.60	16.43	22.01	0.661	0.759	0.853	0.255	0.182	0.121	0.193	0.123	0.067
	DiffusioNeRF (Ours)	Optimized per Scene	16.20	20.34	**25.18**	0.698	0.818	0.883	0.207	0.139	0.095	0.146	0.081	**0.047**

図. LLFF・DTUデータセットで3, 6, 9視点でトレーニングした際の比較結果[1]

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

評価(DTU)

- NeRFによる再構築では光沢あり物質に苦戦するが、反射面やテクスチャ表面は得意

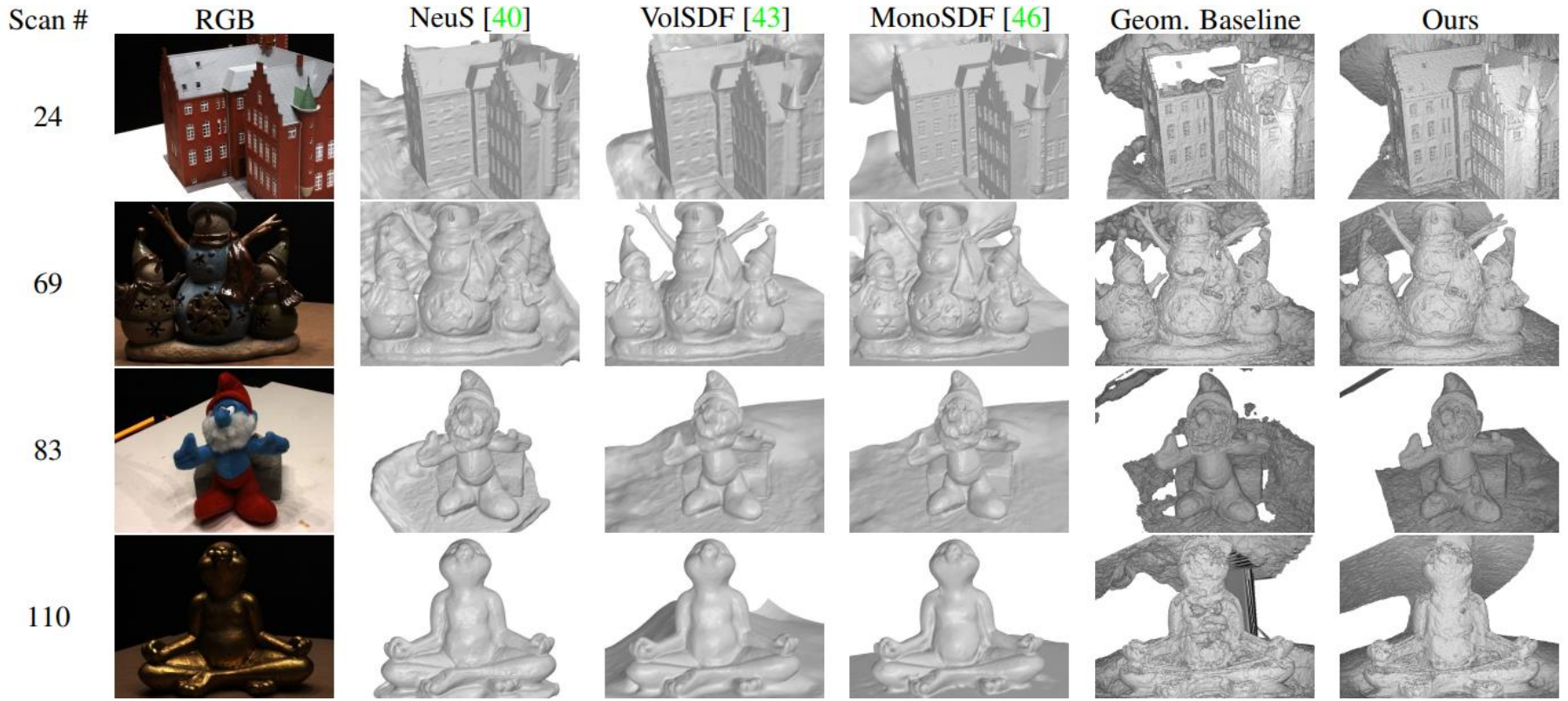

図. DTUデータセットでの評価[1]

論文紹介

DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

評価(Ablation)

- NeRFによる再構築では光沢あり物質に苦戦するが、反射面やテクスチャ表面は得意

Method	LLFF			DTU			
	Average ↓			Average ↓			Chamfer-$L1$ ↓
	3-view	6-view	9-view	3-view	6-view	9-view	All views
$\nabla\mathcal{L} = \nabla\mathcal{L}_{\text{photo}}$	0.210	0.128	0.090	0.203	0.142	0.119	2.87
$\nabla\mathcal{L} = \nabla\mathcal{L}_{\text{photo}} + \lambda_{\text{fg}}\nabla\mathcal{L}_{\text{fg}}$	0.210	0.128	0.090	0.195	0.126	0.092	1.71
$\nabla\mathcal{L} = \nabla\mathcal{L}_{\text{photo}} + \lambda_{\text{fg}}\nabla\mathcal{L}_{\text{fg}} + \lambda_{\text{fr}}\nabla\mathcal{L}_{\text{fr}}$	0.135	0.089	0.072	0.215	0.128	0.093	1.71
$\nabla\mathcal{L} = \nabla\mathcal{L}_{\text{photo}} + \lambda_{\text{fg}}\nabla\mathcal{L}_{\text{fg}} + \lambda_{\text{fr}}\nabla\mathcal{L}_{\text{fr}} - \lambda_{\text{DDM}}\epsilon_\theta$	0.145	0.085	0.066	0.190	0.097	0.072	1.67
$\nabla\mathcal{L} = \nabla\mathcal{L}_{\text{photo}} + \lambda_{\text{fg}}\nabla\mathcal{L}_{\text{fg}} + \lambda_{\text{fr}}\nabla\mathcal{L}_{\text{fr}} + \lambda_{\text{dist}}\nabla\mathcal{L}_{\text{dist}}$	0.118	0.071	0.060	0.185	0.092	0.056	1.36
$\nabla\mathcal{L} = \nabla\mathcal{L}_{\text{photo}} + \lambda_{\text{fg}}\nabla\mathcal{L}_{\text{fg}} + \lambda_{\text{fr}}\nabla\mathcal{L}_{\text{fr}} + \lambda_{\text{dist}}\nabla\mathcal{L}_{\text{dist}} - \lambda_{\text{DDM}}\epsilon_\theta$	0.127	0.075	0.064	0.135	0.052	0.033	1.21
DDM regularizer using 24x24 patches	0.126	0.074	0.061	0.195	0.068	0.043	1.22
24x24 patch DDM & NeRF fitted with $4 \times \lambda_{\text{DDM}}$	0.129	0.074	0.062	0.260	0.080	0.050	1.22
Patches from input images are not given to DDM	0.139	0.078	0.066	0.159	0.063	0.049	1.91
DDM trained with 20% of Hypersim scenes	0.132	0.078	0.066	0.163	0.057	0.035	1.65
RGB-only DDM regularizer	0.134	0.083	0.070	0.189	0.081	0.058	1.31
$\tau = 0$ (no schedule) during NeRF fitting	0.137	0.081	0.067	0.152	0.055	0.042	1.31
NeRF fitted with $4 \times \lambda_{\text{DDM}}$	0.146	0.088	0.076	0.220	0.134	0.071	2.56

図. Ablation studies[1]

[1] Jamie Wynn, Daniyar Turmukhambetov, ” DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models”, CVPR 2023, 2023-11-8

研究方針

やりたいこと

- DiffusionNeRFで3D-Face Reconstructionを行いたい
- 従来のジオメトリ損失関数とDiffusion modelによるスコア関数の他に、損失関数を足したい
  - セマンティックガイダンス(FENeRF)
  - ID Loss(FaceDNeRF)
  - Illumination Loss(FaceDNeRF)
- Diffusion model×テキストプロンプトによる操作を行いたい

課題

- 深度付き顔データが必要(facescape、RGB-D-Face-Database)
- 深度なし顔データセットでも学習できるか？(FFHQ、AFHQv2、CelebAMask-HQ)
- DiffusionNeRFのtiny-cuda-nnの操作性、改変性

図. 作成中のRGB-D Face(facescape由来)

研究方針

データセット

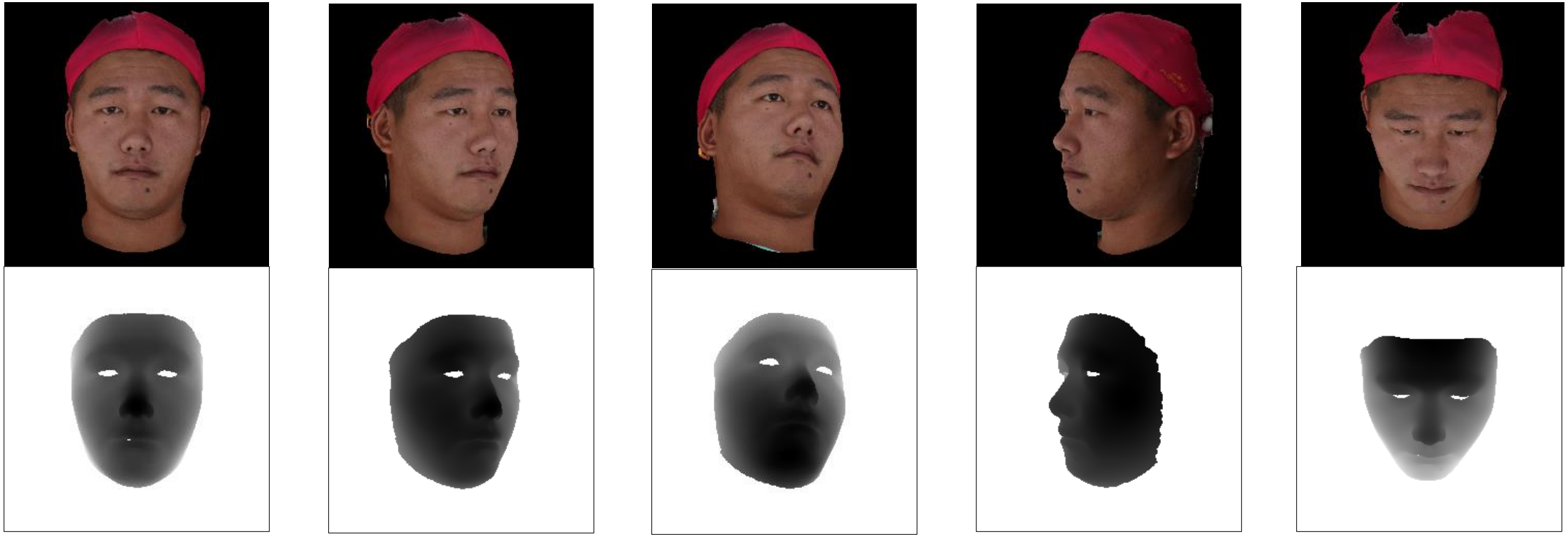

図. 作成中のRGB-D Face(facescape由来)