H-004

サンプリング点の最適化と Coarse-to-fine の統合
Sampling Point Optimization and Coarse-to-fine for Neural Radiance Fields

1）有村玲音　1）太田和宏　1）向田眞志保　1）小野智司
Reo Arimura　Kazuhiro Ohta　Mashiho Mukaida　Satoshi Ono

1　はじめに

Neural Radiance Fields（NeRF）[1] は，対象シーンを複数の異なる視点から撮影した画像セットをもとに，3 次元空間内の各点における色および密度を推定するニューラルネットワークを訓練し，任意の視点からの画像を生成する枠組みである．具体的には，各光線上に配置される複数のサンプリング点の空間座標および視線方向から色および密度を推定し，ボリュームレンダリング [2] を行うことで，最終的な各画素の色を算出する．

一般的な NeRF は Coarse-to-fine（C2F）アプローチにより対象シーン固有の特性をある程度考慮するものの，Coarse 段階において光線上のサンプリング点を一様に配置するため，実際には存在しないノイズであるアーチファクトが発生する，薄い物体の復元に失敗するなどの問題がある．この問題を解決するために，シーンの特性に基づいて適応的にサンプリング点を選択する Sampling Point Optimization（SPO）[3] が提案されている．SPO は，多層パーセプトロン（Multilayer perceptron: MLP）を用いてサンプリング点の位置を最適化し，適応的にサンプリング点を配置することで従来の NeRF と比較して精細な自由視点画像の生成を実現する．

本研究では，SPO に C2F を統合することによる性能の変化を検証する．C2F における Fine 段階のサンプリング点の追加は，ボリュームレンダリング結果における重要度に基づいて行われており，SPO に対しても相補的な効果を持つと考えられ，SPO を Coarse 段階として C2F に統合することで，さらなるレンダリング画像の品質向上が期待できる．

2　関連研究

Kurz らは，レンダリングの効率化を目的とする研究として，AdaNeRF [4] を提案した．AdaNeRF では，サンプリングネットワークを利用し，最終的なピクセルの色に影響を与える可能性の高いサンプリング点を推定する．重要なサンプリング点のみを用いることでレンダリングの効率化を実現している．

著者らは，サンプリング点の重要性に着目してレンダリング画像の品質を改善する SPO [3] を提案した．SPO は，サンプリング点の位置を，光線情報を入力とするサンプリングモジュールにより最適化する．サンプリングモジュールと NeRF モジュールを end-to-end で訓練を行うことが可能である．

3　提案手法

本研究では，従来の NeRF で使用されるサンプリング手法である Coarse-to-fine を SPO に統合することを提案する．SPO のサンプリング手法に，密度情報を用いた Coarse-to-fine によるサンプリング戦略を統合することで，より正確にシーンの特性を考慮した適応的サンプリングが可能である．特に，重要な領域にサンプリング点を集中させることが可能であることから，高精細な自由視点画像が生成可能になることが期待できる．

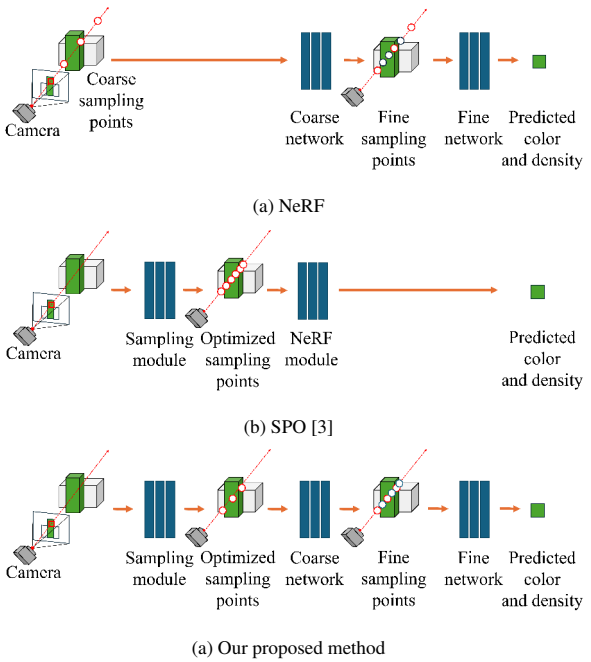


図 1　提案手法の概要．

サンプリングモジュールは，シーン内の N_r 本の光線における光線原点 $\boldsymbol{o}$ と光線方向 $\boldsymbol{d}_j$（$j \in \{1,\ldots,N_r\}$）を入力として，光線上のサンプリング点 i の原点からの距離 $t_{j,i}$（$i \in \{1,\ldots,N_s\}, j \in \{1,\ldots,N_r\}$）を出力する．ここで N_s は 1 本の光線にとるサンプリング点の数を示す．出力されたサンプリング点の距離をもとに，3 次元座標に変換する．（$\mathbf{x}_{j,i} = \mathbf{o} + t_{j,i}\mathbf{d}_j$ ）．

NeRF モジュール F は，カメラ光線の方向ベクトル $\boldsymbol{d}_n$，ならびにサンプリングモジュール G によって出力されたサンプリング点の位置情報 $\boldsymbol{x}_{j,i}$ を入力として，それぞれのサンプリング点における色 $\boldsymbol{c}_{j,i} = (r,g,b)$ と密度 $\sigma_{j,i}$ を出力する．この NeRF モジュール F は，カメラ光線上のサンプリング点 $\boldsymbol{x}_{j,i}$ における色 $\boldsymbol{c}_{j,i}$ と密度 $\sigma_{j,i}$ を学習するシーンに応じて最適化することで，任意の視点から高品質な自由視点画像を生成可能にする．自由視点画像を生成する際には，ボリュームレンダリング [2] を使用して，光線 $\boldsymbol{r}$ に対応する画素の色 $\hat{C}(\boldsymbol{r})$ を算出する．

$$\hat{C}(\boldsymbol{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i\delta_i))\boldsymbol{c}_i \tag{1}$$

ここで，$\delta_i = t_{i+1} - t_i$ は隣接するサンプリング点間の距離である．

一般的な NeRF では Coarse-to-fine アプローチを採用し，パラメータの異なる 2 つの NeRF モジュール $F^{(c)}$ および $F^{(f)}$ を訓練する．Coarse 段階では $N_s/2$ 点のサ

1）　鹿児島大学　Kagoshima University

Algorithm 1 Training

Require: Training dataset $\mathcal{D}$, number of sampling points N_s
Ensure: Trained architecture G, $F^{(c)}$, and $F^{(f)}$
1: Train sampling module G and coarse NeRF module $F^{(c)}$ with $N_s/2$ sampling points.
2: Add fine NeRF module $F^{(f)}$ to the network.
3: Train G, $F^{(c)}$, and $F^{(f)}$ with N_s sampling points.
4: **return** G, $F^{(c)}$, and $F^{(f)}$.

ンプリングを行い $F^{(c)}$ を訓練し，Fine 段階では，$F^{(c)}$ の出力をもとに $N_s/2$ 個のサンプリング点を追加して $F^{(f)}$ を訓練する．提案手法におけるモデルの訓練手順を Algorithm 1 に示す．一般的な NeRF における Coarse 段階においてサンプリングモジュール G と Coarse NeRF モジュール $F^{(c)}$ を訓練し，fine 処理の段階において fine NeRF モジュール $F^{(f)}$ を追加する．Coarse の段階では $N_s/2$ 個のサンプリング点を G により決定し，$F^{(c)}$ は光線当たり $N_s/2$ 個の点における色と密度を推測する．fine の段階では，$F^{(c)}$ により予測された密度に基づいて確率分布を生成し，この分布に従って物体が存在する確率の高い領域に $N_s/2$ 個のサンプリング点を追加する．$\hat{C}(\boldsymbol{r})$ を求めるボリュームレンダリングは得られる色を光線に沿った全てのサンプリング点の色 c_i の加重和として書き換える．

$$\hat{C}_c(\boldsymbol{r}) = \sum_{i=1}^{N_c} w_i c_i\,, \qquad w_i = T_i(1-\exp\left(-\sigma_i\delta_i\right))\,. \tag{2}$$

ここで，w_i は c_i の重みとみなすことができ，w_i が大きいほど，光を遮る物体が存在する可能性が高い．これらの重みを $p_i = w_i/\sum_{j=1}^{N_c} w_j$ により確率密度関数（PDF）として扱うことで，どの区間に物体があるかを表す確率分布となる．重みが大きい区間から多くのサンプリング点を取るために，この PDF の累積分布関数（CDF）$F_i = \sum_{j=1}^{i} p_j$ の逆関数を利用してサンプリングを行う．具体的には，一様乱数 $u \sim \mathcal{U}(0,1)$ を生成し，$F_i \le u \le F_{i+1}$ を満たす i において，以下のような線形補間によって新しいサンプル点 t_i' を算出する．

$$t_i' = t_i + \frac{u - F_i}{F_{i+1} - F_i}(t_{i+1} - t_i) \tag{3}$$

損失関数は従来の NeRF と同様，Coarse ネットワークと fine ネットワークそれぞれが予測したピクセル値とターゲット画像のピクセル値の誤差を最小化する．

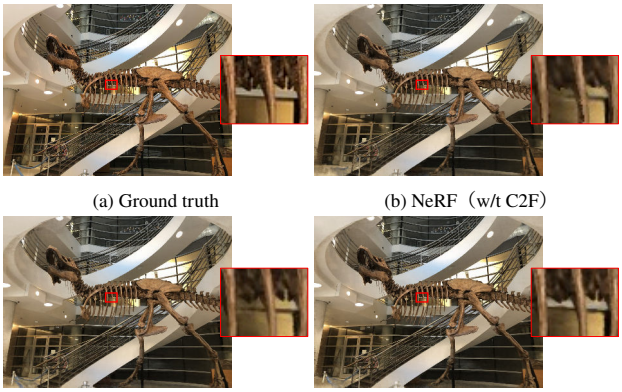
4 評価実験

本研究では，提案手法の有効性を，実画像と COLMAP [5, 6] を用いて推定されたカメラポーズ情報から構成される Real Forward-Facing データセットを用いた実験により検証した．評価指標として，PSNR，SSIM，および LPIPS の 3 つの指標を用いて，生成された画像の品質を調査した．

実験結果を表 1，図 2 に示す．表 1 より，8 シーンの平均を計算した結果 PSNR，SSIM，LPIPS の 3 つの評価指標について品質が向上していることが確認できた．また，図 2 の結果より，T-Rex シーンについて通常の SPO で再現できていない細い骨格を再現できたことが確認できた．

表 1　Real Forward-Facing データセットによる評価結果

	Fern	Flower	Fortress	Horns	Leaves	Orchids	Room	T-Rex	avg
	PSNR↑								
NeRF [1]	24.59	27.18	30.74	26.41	20.66	**20.38**	31.47	25.68	25.89
SPO [3]	24.97	27.62	30.77	26.79	20.92	20.10	32.25	26.71	26.27
Ours	**25.19**	**27.70**	**31.28**	**27.47**	**21.00**	20.31	**32.86**	**27.12**	**26.62**
	SSIM↑								
NeRF [1]	0.734	0.792	0.859	0.753	0.632	0.574	0.909	0.822	0.759
SPO [3]	0.757	0.807	0.859	0.773	0.655	0.568	0.920	0.852	0.774
Ours	**0.767**	**0.812**	**0.881**	**0.801**	**0.662**	**0.587**	**0.925**	**0.864**	**0.787**
	Lpips↓								
NeRF [1]	0.323	0.238	0.201	0.325	0.344	0.341	0.209	0.290	0.284
SPO [3]	0.292	0.220	0.200	0.296	0.318	0.339	0.179	0.251	0.262
Ours	**0.279**	**0.213**	**0.161**	**0.263**	**0.309**	**0.325**	**0.165**	**0.238**	**0.244**



(a) Ground truth　(b) NeRF（w/t C2F）
(c) SPO（w/o C2F）　(d) SPO（w/t C2F）

図 2　Real Forward-Facing データセットによる実験結果

5 結論

本研究では，生成画像の品質向上を目的として，従来の NeRF で使用されるサンプリング手法である Coarse-to-fine を SPO に統合することを提案した．実験の結果，SPO に Coarse-to-fine を統合することでよりシーンの特性を考慮した適応的にサンプリング点の位置を最適化することが可能であり，生成画像の品質が向上したことを確認した．今後，さらなるレンダリング品質の改善を目指す．

参考文献

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[2] J. T. Kajiya *et al.*, "Ray tracing volume densities," *ACM SIGGRAPH computer graphics*, vol. 18, no. 3, pp. 165–174, 1984.

[3] K. Ohta and S. Ono, "Neural radiance field image refinement through end-to-end sampling point optimization," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. n/a, no. n/a, 2024. DOI: https://doi.org/10.1002/tee.24222.

[4] A. Kurz *et al.*, "Adanerf: Adaptive sampling for real-time rendering of neural radiance fields," in *European Conference on Computer Vision*, pp. 254–270, Springer, 2022.

[5] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4104–4113, 2016.

[6] J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, "Pixelwise view selection for unstructured multi-view stereo," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pp. 501–518, Springer, 2016.