情報処理学会研究報告
IPSJ SIG Technical Report

多層パーセプトロンと3D Gaussian Splattingを用いたフォーカス位置可変なシーン表現

簸本 悠紀[1]　西田 拓央[1]　井尻 敬[1]

概要：本研究の目的は，自然な焦点ボケを含む画像をレンダリングでき，かつ，レンダリング時に自由にフォーカス位置を変更可能な 3DGS によるシーン表現手法の確立である．このため本研究では，3D Gaussian Splatting に多層パーセプトロン（MLP）を組み合わせた手法を提案する．具体的には，各 Gaussian の位置・視点から各 Gaussian への方向・フォーカス位置を考慮しながら，MLP を用いてそのスケール・色・透明度を決定する．提案手法の有用性を確認するため，複数の 3 次元シーンの再構成を実施し，レンダリング精度を PSNR・SSIM・LPIPS により評価し，その結果を示す．

図 1: 提案手法を用いてフォーカスパラメータ $f \in [0, 1]$ を変化させながらレンダリングした結果.

1. はじめに

フォーカスブラケット撮影とは，フォーカス位置を前後に移動しながら複数枚の写真を撮影する手法である．深度合成とは，フォーカスブラケット撮影により得られた一連の写真から，被写界深度の深い写真を合成する手法である．これらは，被写界深度が浅いマクロレンズを使用する必要のある，花や昆虫標本などの微小物体の写真撮影の際に用いられる技術である．また，複数視点よりフォーカスブラケット撮影を実施し，各視点にて深度合成写真を作成した後，フォトグラメトリを適用することで，微小物体の 3 次元モデルを構築する手法が提案されている [3, 13, 14]．しかし，フォトグラメトリには，昆虫標本の翅や目などの透明部分の形状復元が難しいという課題や，再構成された 3 次元モデルには実際の写真に見られる自然な焦点ボケが失われるという課題がある．

複数視点から撮影された写真を入力とし，各写真の見た目を再現する新たな 3 次元シーンの表現手法である 3D Gaussian Splatting（3DGS）[5] が，注目を集めている．この手法は，3 次元空間上に色・透明度・スケール・回転のパラメータを持つ Gaussian を配置することで 3 次元シーンを表現する．この手法では，多視点写真から高精度なシーン表現の高速な学習が可能であり，フォトグラメトリでは難しい半透明物体の表現が可能である．しかし，3DGS は焦点ボケのない全焦点写真の入力を仮定しているため，自然なボケを含むシーンのレンダリングは行えない．

本研究の目的は，自然な焦点ボケを含む画像をレンダリングでき，かつ，レンダリング時に自由にフォーカス

[1] 芝浦工業大学

位置を変更可能な 3DGS によるシーン表現手法の確立である．このため本研究では，3DGS に多層パーセプトロン（MLP）を組み合わせた手法を提案する．提案手法では，各 Gaussian の位置・視点から各 Gaussian への方向・フォーカスパラメータを入力し，その Gaussian のスケール・色・透明度の係数を出力する MLP を用いたモデルを利用する．提案手法の入力は多視点フォーカスブラケット写真と多視点深度合成写真である．提案手法は，最初に多視点深度合成写真群を用いてボケのない 3 次元シーン表現を学習し，続いて多視点フォーカスブラケット写真を用いてモデルを学習することでフォーカス位置を変更できる表現を学習する．図 1，5 に，提案手法を用いたレンダリング結果を示す．フォーカスパラメータ f を指定することで自由にフォーカス位置を変更でき，かつ，自然な焦点ボケを再現できていることがわかる．

2. 関連研究

2.1 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [5] は, 3 次元の Gaussian を空間に散りばめることで 3 次元シーンを表現する手法である．シーンの構成要素となる各 Gaussian は次の通り表現される，

$$G(\mathbf{x}) = e^{-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}}, \tag{1}$$

$$\mathbf{\Sigma} = \mathbf{RSS}^T\mathbf{R}^T. \tag{2}$$

ただし，$\mathbf{x}$ は中心の座標，$\mathbf{\Sigma}$ は分散共分散行列，$\mathbf{R}$ は回転行列，$\mathbf{S}$ は異方性を持つスケール行列である．加えて，各 Gaussian は，視線方向に依存する色 $\mathbf{c} \in \mathbb{R}^3$ と透明度 $\alpha \in \mathbb{R}$ の属性を持つ．なお，$\mathbf{c}$ は球面調和関数により表現される．これらの Gaussian の位置・大きさ・見た目を決定する属性は，多視点写真を用いた学習により決定される．また，レンダリングは，全 Gaussian を手前のものからスクリーンへ投影することで実施される．

この 3DGS では，入力された多視点写真を再現するように各 Gaussian の属性を最適化するため，多様な反射特性を含むシーンを表現可能である．しかし，一般的に，3DGS は焦点ボケを含まない入力画像を仮定しており，焦点ボケを含むシーン表現や，レンダリング時にフォーカス位置を変更するような表現は行えない．

2.2 ライトフィールドを用いたフォーカス位置可変なシーン表現

ライトフィールドとは，空間を通過する光の位置と方向を記録するシーン表現方法であり [7]，ライトフィールドカメラを用いることで計測できる [12]．従来のカメラはレンズを通ってセンサに届く光の総量を 2 次元の写真として計測するのに対して，ライトフィールドカメラは主レンズと撮像センサの間にマイクロレンズアレイを配置することで，センサに到達する光の位置と方向を 4 次元データとして記録する．このライトフィールドを用いることで，レンダリング時に，フォーカス位置や絞りを変更した画像を合成できる．しかし，この手法は 3 次元シーンを構築するものでないため，視点位置を大きく変更した画像合成は行えない．

2.3 NeRF を用いたフォーカス位置可変なシーン表現

Neural Radiance Field（NeRF）[9] は，3 次元シーン内の位置と視線方向から，その点の密度と色を出力する深層学習モデル $\mathcal{H}_\Theta$ を用いて 3 次元シーンを表現する．この $\mathcal{H}_\Theta$ は，密度を推定する Density MLP と，色を推定する Color MLP という 2 つの MLP から構成される．レンダリングは，視点からスクリーン上の各ピクセルに飛ばしたレイ上のサンプリング点において $\mathcal{H}_\Theta$ を評価し，出力される各点の密度を考慮して色を混合することで行われる．

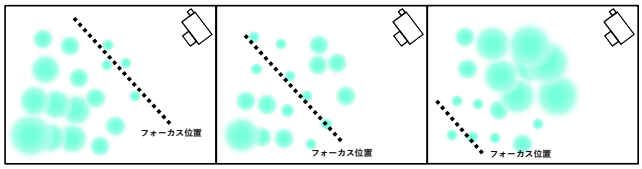
NeRF を応用したフォーカス位置可変なシーン表現手法が研究されている．DoF-NeRF [19] は，通常の NeRF と同じシーン表現を利用し，レンダリング時に工夫を施すことで，焦点ボケの効果を追加する．具体的には，レンダリングの際にユーザより入力される「焦点距離」と「カメラの絞り」という 2 つのパラメータを考慮してサンプリング点を発散させることで，焦点ボケの効果を付与する．ただし，色の発散を考慮する点をレイに対して 1 つの点に近似しているため，高精度なボケの再現が難しい．一方，Refocus-NeRF [20] は，シーンを表現するモデル自体を焦点ボケ効果を含むように拡張した．具体的には，シーンを表現するモデル $\mathcal{H}_\Theta$ を，フォーカス位置を表すパラメータも入力できるように拡張し，多視点フォーカスブラケット撮影した写真群によりこのモデルを訓練した．しかし，NeRF ベースのシーン表現手法であるためレンダリングコストが非常に高く，FPS が 20 程度とリアルタイム性に欠ける．

3. フォーカス位置可変なシーン表現

本研究では，3DGS をフォーカス位置可変となるように拡張した新たなシーン表現手法を提案する．まず，提案するシーン表現について解説した後（3.1 節），Gaussian のスケール・色・透明度を決定する MLP を用いたモデルについて説明する（3.2 節）．その後，次章にて多視点フォーカスブラケット撮影を用いたシーン表現の学習方法について詳しく解説する．

3.1 3DGS を拡張したフォーカス位置可変なシーン表現

本研究では，3DGS を基礎として，フォーカス位置可変な新たなシーン表現手法を提案する．図 2 にシーン表現の概略を示す．提案手法は，3DGS と同様に，Gaussian を 3 次元空間上に配置することでシーンを表現する．各 Gaussian



$f = 0.0$ $f = 0.5$ $f = 1.0$

図 2: 3D Gaussian Splatting に基づくフォーカス位置可変なシーン表現．フォーカスパラメータ $f \in [0, 1]$ によって各 Gaussian のスケール・色・透明度を変化させることで，自然な焦点ボケを再現する．

は，位置 $\mathbf{x} \in \mathbb{R}^3$，回転 $\mathbf{R} \in \mathbb{R}^3$，スケール $\mathbf{S} \in \mathbb{R}^3$，色 $\mathbf{c} \in \mathbb{R}^{d_c}$．透明度 $\alpha \in \mathbb{R}$，という 5 種の属性により定義される．ここで，『色 $\mathbf{c}$』は視線方向に依存する色であり，RGB 各チャンネルごとに 3 次の球面調和関数を用いて表現される．そのため，$\mathbf{c}$ の次元は $d_c = 48$ (3 チャンネル $\times$3 次球面調和関数の係数 16 個) となる．提案手法では，レンダリングの際に，視点位置に加えてフォーカスパラメータ $f \in [0, 1]$ が与えられるものとする．我々は，この f に応じて，各 Gaussian のスケール $\mathbf{S}$・色 $\mathbf{c}$・透明度 α を変化させることで，焦点ボケの効果を再現する．

ここで，Gaussian の $(\mathbf{S}, \mathbf{c}, \alpha)$ の変化量は，シーンの構造や特徴に依存して決定する必要がある．例えば，視点から各 Gaussian への深度と，フォーカス位置のずれを利用して Gaussian の属性を変化させるような単純な手法は，シーン内に鏡のような鏡面反射する物体が含まれる場合に正確に機能しない．そこで本研究では，各 Gaussian の $(\mathbf{S}, \mathbf{c}, \alpha)$ の変化量を MLP を用いたモデルにより決定する．提案手法では，多視点深度合成写真を用いた Gaussian によるの 3 次元シーン再構成後に，フォーカスブラケット写真を用いてモデルを最適化する．

3.2 MLP を用いたモデルによる Gaussian の属性変化

我々は，シーン全体に対して MLP を用いたリフォーカスモデル $\mathcal{F}_\Theta$ を用意し，これを利用してシーン内のすべての Gaussian のスケール・色・透明度の変化量を計算する，

$$\mathcal{F}_\Theta : (\mathbf{x}, \mathbf{d}, f) \to (\gamma_{\text{scale}}, \gamma_{\mathbf{c}}, \gamma_\alpha). \tag{3}$$

ここで，$\mathbf{x}$ は Gaussian の中心座標，$\mathbf{d}$ は Gaussian への視線方向ベクトル，f はフォーカス位置を表すパラメータである．また，$(\gamma_{\text{scale}} \in \mathbb{R}^3_+, \gamma_{\mathbf{c}} \in \mathbb{R}^{d_c}_+, \gamma_\alpha \in \mathbb{R}_+)$ は，それぞれ，Gaussian の $(\mathbf{S}, \mathbf{c}, \alpha)$ に掛ける 0 以上の係数である．

我々は，リフォーカスモデル $\mathcal{F}_\Theta$ を，2 つの MLP により構成する（図 3）．一つは『透明度係数 MLP』で，$\mathbf{x}$ と f を入力として受け取り，γ_α と特徴ベクトルを出力する．もう一つは『スケール・色係数 MLP』で，特徴ベクトルと $\mathbf{d}$ を入力として受け取り，γ_{scale} と $\gamma_{\mathbf{c}}$ を出力する．各 MLP は 64 個のニューロンを持つ 3 層の中間層からなり，活性化関数には ReLU 関数を用いる．透明度係数 MLP に $\mathbf{d}$ を入力しない理由は，NeRF [9] と同様に，Gaussian の透明度が視線方向に依存して変化しないようにするためである．

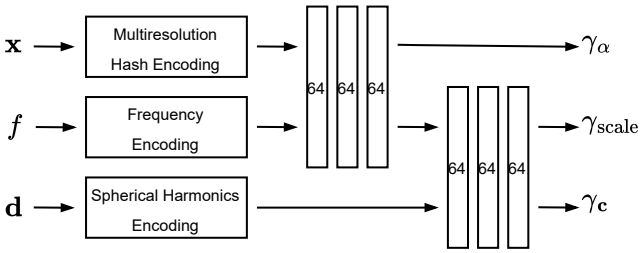


図 3: MLP を用いたリフォーカスモデル $\mathcal{F}_\Theta$ の構造．

本研究では，$\mathcal{F}_\Theta$ への各入力に対して，高次元ベクトルに変換するエンコードを施す．まず，3 次元座標 $\mathbf{x}$ には，Multiresolution Hash Encoding [11] を適用し，$\mathcal{F}_\Theta$ とともに特徴ベクトルを含むハッシュテーブルを学習する．次に，視点方向ベクトル $\mathbf{d}$ には Spherieal Harmonics Encoding [2, 11, 17] を適用する．最後に，フォーカスパラメータ f には Frequency Encoding [9, 16],

$$\begin{aligned}\text{enc}(f) = (&\sin(2^0 f), \sin(2^1 f), \ldots, \sin(2^{L-1} f),\\ &\cos(2^0 f), \cos(2^1 f), \ldots, \cos(2^{L-1} f))\end{aligned} \tag{4}$$

を適用する．また，学習と推論の高速化のため，我々は，$\mathcal{F}_\Theta$ の実装に tiny-cuda-nn [10] を使用する．

4. 多視点写真を利用した 3D シーンの構築法

4.1 多視点フォーカスブラケット撮影による入力データの準備

前章にて紹介した 3D シーンを構築するため，まず我々は，複数視点よりフォーカスブラケット撮影を実施する (図 4a)．続いて，各視点において撮影されたフォーカス位置の異なる一連の写真群について，フォーカス位置の変化に伴う画角変化（フォーカスブリージング）を補正するアラインメント処理を施す（図 4b）．続いて，アラインメント処理された写真にピラミッド法を適用することで，写真全体にピントの合った深度合成写真を取得する（図 4c）．

なお，我々は，各視点におけるアラインメントされた一連の写真群（図 4b）について，i 番目の写真のフォーカスパラメータを $f_i = i/(N-1)$ と定義する．ただし，N はフォーカスブラケット撮影にて撮影した写真の枚数であり，フォーカス位置が最も手前の写真が $f_0 = 0$ に対応する．

最後に，異なる視点から撮影された深度合成写真に，Structure from Motion (SfM) と Multi-View Stereo (MVS) を適用することで，カメラ位置と 3 次元点群を取得する (図 4d)．本研究では，アラインメント処理と深度合成には Helicon Focus を，SfM と MVS には Metashape を利用した．

4.2 シーン最適化の流れ

本研究では，次の 3 ステップによりシーンを構築する

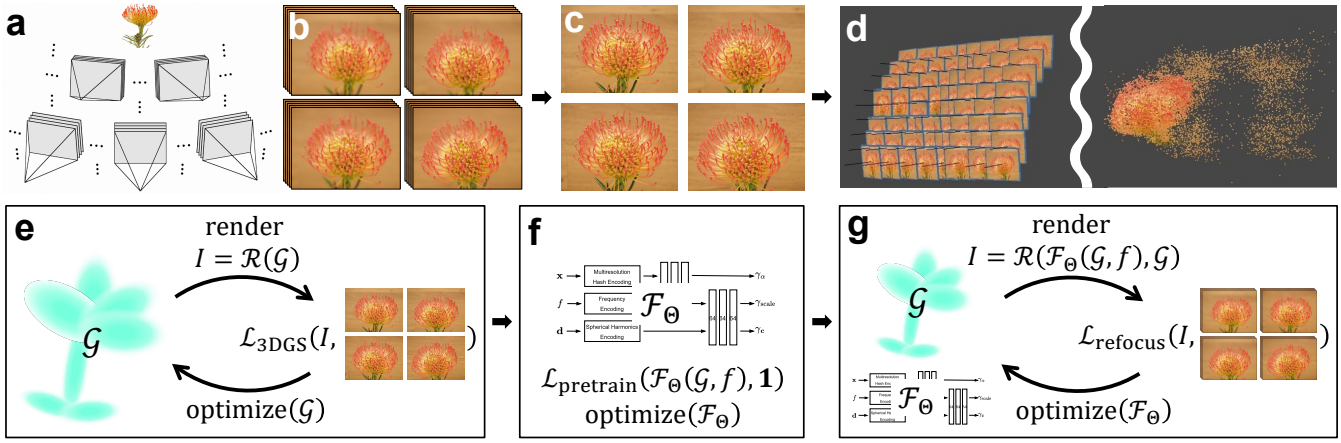


図 4: 3 次元シーン復元の流れ．(a) 複数視点からフォーカスブラケット撮影を実施し，(b) 各視点におけるフォーカス位置の異なる一連の写真をアラインメントした後，(c) 各視点における一連の写真に深度合成を適用する．(d) その後，多視点深度合成写真に SfM と MVS を適用しカメラ位置と点群を推定する．シーン復元は，(e) Step 1：シーンを構成する Gaussian 群 $\mathcal{G}$ の属性の最適化，(f) Step 2：$\mathcal{F}_\Theta$ の事前学習，および，(g) Step 3：$\mathcal{F}_\Theta$ の最適化，より構成される．

Gaussian 群 $\mathcal{G}$ の属性と，リフォーカスモデル $\mathcal{F}_\Theta$ を最適化する．最初のステップでは，$\mathcal{F}_\Theta$ の出力を $(1, 1, 1)$ に固定し，深度合成写真を用いて $\mathcal{G}$ の属性を最適化する．これにより，焦点ボケを含まないシーンが構築される．これ以降，$\mathcal{G}$ の属性は固定し，Gaussian の追加・削除の処理を行わない．次のステップにて，リフォーカスモデル $\mathcal{F}_\Theta$ の事前学習を実施する．最後のステップにおいて，多視点フォーカスブラケット写真を用いて $\mathcal{F}_\Theta$ を訓練することで，焦点ボケを含むシーンを構築する．

4.3 Step 1. Gaussian 属性の最適化

Step 1 では，複数視点より撮影された深度合成写真群を用いてシーンを構成する Gaussian 群 $\mathcal{G}$ の属性を最適化する（図 4e）．まず，SfM および MVS により作成した 3 次元点群を用いて Gaussian を初期化する．次に，深度合成写真群を目標画像とし，シーンを構成する Gaussian 群 $\mathcal{G}$ を用いたレンダリング結果 $\mathcal{R}(\mathcal{G})$ との損失を最小化するように，各 Gaussian の属性を逐次的に更新する．

最適化の損失関数には，3DGS [5] にて提案された $\mathcal{L}_{\text{3DGS}} = (1-\lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{\text{D-SSIM}}$ を利用する．また，学習率などのパラメータにも 3DGS と同様のものを設定し，30k 回の逐次最適化を実施する．

4.4 Step 2. リフォーカスモデル $\mathcal{F}_\Theta$ の事前学習

Step 2 では，リフォーカスモデル $\mathcal{F}_\Theta$ が不正確な局所解に陥るの防ぐために事前学習を行う（図 4f）．具体的には，シーンを構成する Gaussian 群 $\mathcal{G}$ に対して，リフォーカスモデル $\mathcal{F}_\Theta$ が各係数 $(\gamma_{\text{scale}}, \gamma_{\mathbf{c}}, \gamma_\alpha)$ について 1 を出力するように学習する．そこで，損失関数には平均二乗誤差を使用する，

$$\mathcal{L}_{\text{pretrain}} = \frac{1}{|\mathcal{E}||\mathcal{G}|N}\sum_{e\in\mathcal{E}}\sum_{g\in\mathcal{G}}\sum_{i=0}^{N-1}||\mathcal{F}_\Theta(g.\mathbf{x}, e.\mathbf{d}, f_i) - (1,1,1)||^2. \tag{5}$$

ただし，$\mathcal{E}$ は SfM により計算されたカメラ位置の集合であり，f_i は前から i 番目のフォーカスブラケット写真に対応付けられたフォーカスパラメータである．

本研究では，バッチサイズを N として，Adam [6]（学習率 10^{-3}）を使用して 1,000 イテレーション学習する．ただし，全バッチを回る毎（1 エポック毎）に，学習率を 0.8 倍減衰させる．また，学習の高速化のために混合精度学習を行う [8]．

4.5 Step 3. リフォーカスモデル $\mathcal{F}_\Theta$ の最適化

Step 3 では，焦点ボケを含むフォーカスブラケット写真群を正解画像とし，リフォーカスモデルを $\mathcal{F}_\Theta$ 学習する（図 4g）．視点 $e \in \mathcal{E}$ から撮影された一連の写真のうち，フォーカス位置が前から i 番目のものを I_i^e とする．また，リフォーカスモデル $\mathcal{F}_\Theta$ に，Gaussian 位置 $\mathbf{x}$・視点 e から Gaussian への視線方向 $\mathbf{d}$・フォーカスパラメータ f_i を入力して得られる係数を用いて，シーンを構成する Gaussian 群 $\mathcal{G}$ を変換した結果を $\mathcal{G}'$ とする．この $\mathcal{G}'$ をレンダリングした画像 $\mathcal{R}(\mathcal{G}')$ と，正解画像 I_i^e との損失を計算する．本研究では，VGG 損失 [1, 4, 15] と LPIPS 指標 [21] を組み合わせた以下の損失関数を利用する，

$$\mathcal{L}_{\text{refocus}} = \mathcal{L}_{\text{VGG}} + 0.5\mathcal{L}_{\text{LPIPS}}. \tag{6}$$

学習の際，バッチサイズは N とし，1 つのバッチに 1 視点分の N 枚のフォーカスブラケット写真群が含まれるようにする．学習には Adam（学習率 10^{-4}）を使用し，5,000

表 1: 各データセットにおける PSNR，SSIM [18]，LPIPS [21] 指標による評価結果の平均値.

データセット	PSNR	SSIM	LPIPS
ピンクッション（図 1）	32.3	0.994	0.060
セロシア（図 5A）	30.1	0.976	0.063
ソラナムパンプキン（図 5B）	33.2	0.997	0.046
ナノブロック（図 5C）	29.1	0.991	0.058
ホウセキゾウムシ（図 5D）	34.5	0.994	0.014

表 2: 各データセットにおける，フレームレート（FPS）の平均値とシーンを構成する Gaussian の数（#Gaussians）.

データセット	FPS	#Gaussians
ピンクッション（図 1）	81.4	1,022,764
セロシア（図 5A）	54.4	1,742,474
ソラナムパンプキン（図 5B）	81.5	1,107,490
ナノブロック（図 5C）	45.8	2,062,397
ホウセキゾウムシ（図 5D）	172.8	381,467

イテレーション学習する．ただし，800 イテレーション以降は，全バッチを回る毎（1 エポック毎）に学習率を 0.9 倍減衰させる．また，リフォーカスモデルの事前学習と同様に混合精度学習を行う．

5. 評価実験

5.1 実験の詳細

提案手法の有用性を確認するために，提案手法を用いて複数の異なるシーンの再構成を行った．昆虫標本や植物，ナノブロックなどの被写体を，正面方向から複数視点より撮影した（forward facing シーン）．撮影には自動フォーカスブラケット機能を搭載した OLYMPUS E-M1 Mark II 使用し，撮影視点数は被写体に応じて 42∼64 視点とした．フォーカスブラケット撮影では，フォーカスの移動範囲が，被写体の手前から奥までをカバーするように設定した．1 視点あたりの撮影数 N はシーンに応じて変化し，10∼20 であった．

各シーンを撮影したデータセットにおいて，ランダムに選択した 3 視点分のフォーカスブラケット写真群と深度合成写真群を学習に利用しないテスト用データとし，残りのデータを利用して提案手法によるシーン再構成を実施した．再構成したシーンを用いて，テスト用視点からレンダリング結果を作成し，正解画像と比較した．比較指標には，PSNR，SSIM [18]，LPIPS [21] を用いた．なお，Gaussian 属性の最適化とリフォーカスモデル $\mathcal{F}_\Theta$ の学習には，NVIDIA GeForce RTX3090 を使用した．

5.2 結果と考察

図 1，5 に，テスト用視点における，フォーカスブラケット写真と提案手法によるレンダリング結果を示す．図 1，5 の左端（$f = 0.0$）を見ると，被写体手前にフォーカスが合い，被写体の奥や背景に焦点ボケの効果が現れていることがわかる．提案手法は，リフォーカスモデル $\mathcal{F}_\Theta$ を導入し Gaussian のスケール・色・透明度を変化させることで，3DGS の枠組みに自然な焦点ボケの効果を組み込めた事がわかる．また，図 1，5 の左端から右端に掛けてフォーカスパラメータ f を変化させたレンダリング結果を示す．提案手法を用いると，このように，フォーカスパラメータ f を指定することで，シーン復元後にフォーカス位置を移動させることが可能となる．

表 1 に各指標による評価結果の平均値を示す．全てのシーンにおいて PSNR は 29.0 以上，SSIM は 0.990 以上，LPIPS は 0.070 以下と高い精度が得られた．また，表 2 に各シーンをレンダリングした際のフレームレート（FPS）の平均値とそのシーンを構成する Gaussian の数を示す．全てのシーンにおいて 40.0FPS 以上のレンダリングを達成した．提案手法は 3DGS に基づくため，NeRF に基づく手法 [19,20] と比較し，高速なレンダリングが可能である．また，シーンを構成する Gaussian の数に比例して FPS が変化するが，提案手法ではスクリーンに投影する各 Gaussian にリフォーカスモデル $\mathcal{F}_\Theta$ を適用するのみであるため，提案手法の計算複雑度は 3DGS のレンダリングアルゴリズムと同等である．

6. まとめと展望

本研究では，レンダリング時にフォーカス位置を自由に変更可能な 3 次元シーン表現法の実現を目的とし，3DGS を拡張した手法を提案した．鍵となるアイディアは，入力されたフォーカスパラメータ f に応じて，シーンを構成する Gaussian の属性の変化量を決定する MLP を用いたリフォーカスモデル $\mathcal{F}_\Theta$ の導入である．提案手法を用いて複数のシーンを再構成したところ，提案手法は，リアルタイムかつ高精度に自然な焦点ボケを含むレンダリング画像を出力できることを確認した．

提案手法の課題は，リフォーカスモデルを学習する際，シーンを構成する Gaussian の数に比例する GPU メモリ容量を必要とする点にある．このため，配置可能な Gaussian 数が GPU メモリの大きさにより制限されてしまう．今後，レンダリング結果に寄与しない Gaussian のトリミングを行うなどにより，学習時の GPU メモリ使用量の削減法の開発を行いたい．また，多視点フォーカスブラケット写真ではなく，単純な多視点写真からリフォーカス可能なシーン表現を学習する手法の確立も，重要な将来課題の 1 つである．

謝辞 本研究は JSPS 科研費 23K24965 の助成を受けて実施されたものである．

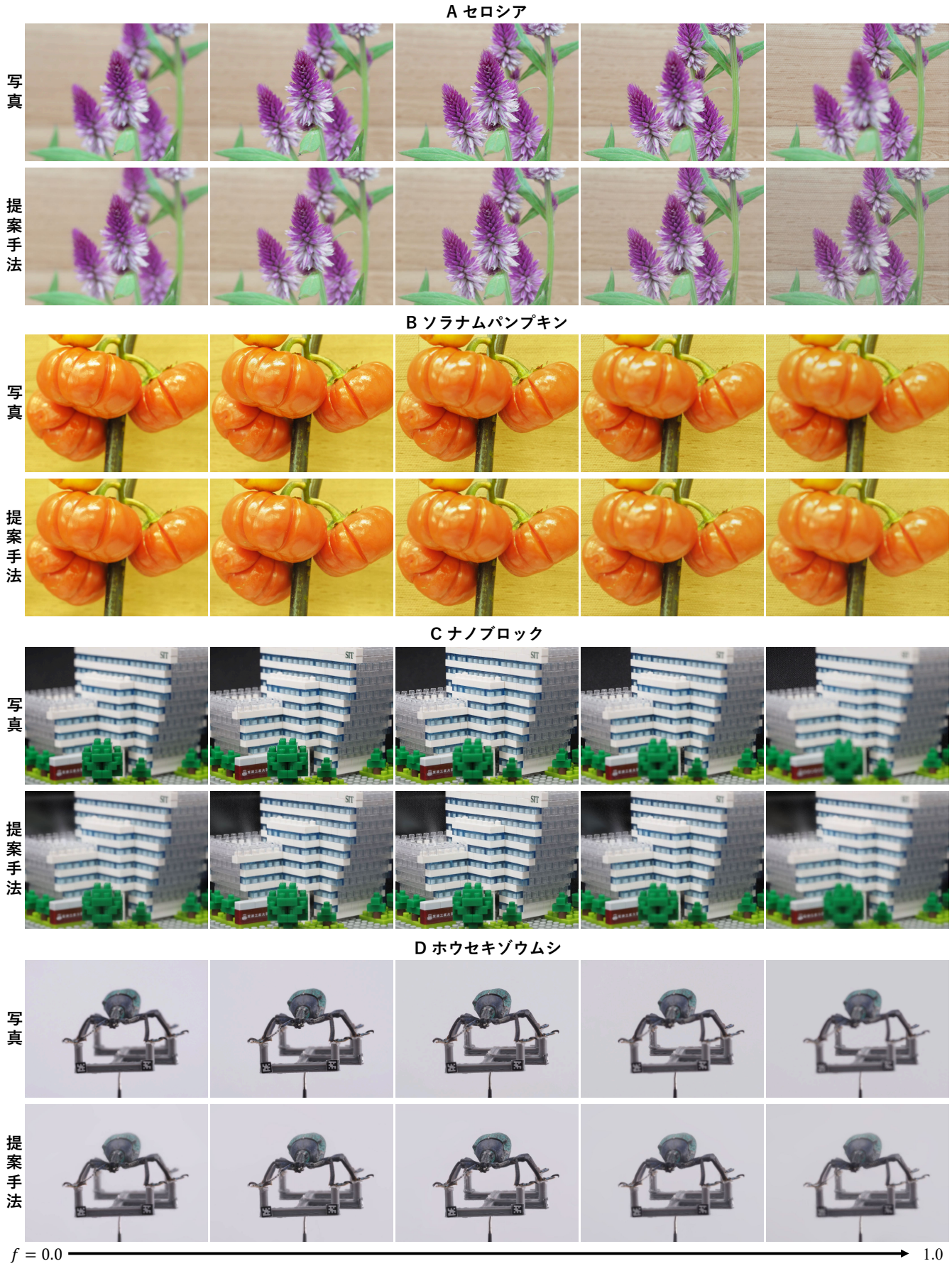


図 5: 4 つのシーンにおけるフォーカスブラケット写真と提案手法によるレンダリング結果の比較.

参考文献

[1] Franke, L., Rückert, D., Fink, L. and Stamminger, M.: TRIPS: Trilinear Point Splatting for Real-Time Radiance Field Rendering, *Computer Graphics Forum*, Vol. 43, No. 2 (2024).

[2] Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B. and Kanazawa, A.: Plenoxels: Radiance Fields Without Neural Networks, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5501–5510 (2022).

[3] Gallo, A., Muzzupappa, M. and Bruno, F.: 3D reconstruction of small sized objects from a sequence of multi-focused images, *Journal of Cultural Heritage*, Vol. 15, No. 2, pp. 173–182 (2014).

[4] Johnson, J., Alahi, A. and Fei-Fei, L.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution, *Proceedings of the European Conference on Computer Vision (ECCV)* (2016).

[5] Kerbl, B., Kopanas, G., Leimkühler, T. and Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering, *ACM Transactions on Graphics*, Vol. 42, No. 4 (2023).

[6] Kingma, D. P. and Ba, J. L.: Adam: A method for stochastic optimization, *Proceedings of the International Conference on Learning Representations (ICLR)* (2015).

[7] Levoy, M. and Hanrahan, P.: Light field rendering, *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, New York, NY, USA, Association for Computing Machinery, p. 31–42 (1996).

[8] Micikevicius, P., Narang, S., Alben, J., Diamos, G. F., Elsen, E., García, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G. and Wu, H.: Mixed Precision Training, *Proceedings of the International Conference on Learning Representations (ICLR)* (2018).

[9] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R. and Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, *Proceedings of the European Conference on Computer Vision (ECCV)* (2020).

[10] Müller, T.: tiny-cuda-nn (2021).

[11] Müller, T., Evans, A., Schied, C. and Keller, A.: Instant Neural Graphics Primitives with a Multiresolution Hash Encoding, *ACM Transactions on Graphics*, Vol. 41, No. 4, pp. 102:1–102:15 (2022).

[12] Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M. and Hanrahan, P.: Light Field Photography with a Hand-held Plenoptic Camera, Research Report CSTR 2005-02, Stanford university (2005).

[13] Nguyen, C. V., Lovell, D. R., Adcock, M. and La Salle, J.: Capturing Natural-Colour 3D Models of Insects for Species Discovery and Diagnostics, *PLOS ONE*, Vol. 9, No. 4, pp. 1–11 (2014).

[14] Qiu, Y., Inagaki, D., Kohiyama, K., Tanaka, H. and Ijiri, T.: Focus stacking by multi-viewpoint focus bracketing, *Proceedings of the SIGGRAPH Asia 2019 Posters*, SA '19, New York, NY, USA, Association for Computing Machinery (2019).

[15] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *Proceedings of the International Conference on Learning Representations (ICLR)* (2015).

[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u. and Polosukhin, I.: Attention is All you Need, *Proceedings of the Advances in Neural Information Processing Systems* (Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R., eds.), Vol. 30, Curran Associates, Inc. (2017).

[17] Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J. T. and Srinivasan, P. P.: Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5481–5490 (2022).

[18] Wang, Z., Bovik, A., Sheikh, H. and Simoncelli, E.: Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612 (2004).

[19] Wu, Z., Li, X., Peng, J., Lu, H., Cao, Z. and Zhong, W.: DoF-NeRF: Depth-of-Field Meets Neural Radiance Fields, *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, New York, NY, USA, Association for Computing Machinery, p. 1718–1729 (2022).

[20] Yabumoto, Y., Nishida, T. and Ijiri, T.: Refocus-NeRF: Focus-Distance-Aware Neural Radiance Fields Trained with Focus Bracket Photography, *Proceedings of the 20th ACM SIGGRAPH European Conference on Visual Media Production* (2023).

[21] Zhang, R., Isola, P., Efros, A. A., Shechtman, E. and Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).