

# 拡散モデルとニューラル場による 共焦点画像の高速な奥行補間

大橋 智哉<sup>1,a)</sup> 藤村 友貴<sup>1</sup> 船富 卓哉<sup>1</sup> 向川 康博<sup>1</sup> Nath Sristilekha<sup>2</sup> 奥田 覚<sup>2</sup>

**概要：**本研究では、拡散モデルとニューラル場を用いた共焦点画像の高速な奥行補間を提案する。共焦点顕微鏡を用いて3次元空間を断層的に撮影した共焦点画像は、奥行方向の撮影枚数が少ない場合は撮影対象の観察は困難となる。この問題に対して、拡散モデルを用いて奥行方向の画像を生成する手法が存在するが、高解像度の画像生成は計算コストが高いという問題がある。したがって本研究では、拡散モデルを用いて生成した低解像度の画像で奥行方向に粗く補間し、さらにニューラル場をフィッティングさせることで、奥行と画像平面の両方向に高解像度な共焦点画像を出力する手法を提案する。提案手法では、拡散モデル単体の場合と比較して、計算時間を大幅に削減しながら、高品質な共焦点画像の生成を実現した。

**キーワード：** Diffusion models, neural fields, confocal microscopy images, depth interpolation

## 1. はじめに

近年、生体組織や細胞の3次元構造を観察するために、共焦点顕微鏡が広く利用されている。共焦点顕微鏡は対象物を断層的に撮影することで高精度な3次元イメージングを可能とし、生物学や医学の研究において重要な役割を果たしている。

共焦点顕微鏡による立体構造の詳細な観察においては、奥行方向に密な撮影をした共焦点画像が望まれる。しかしながら、そのような密な撮影は計測時間の増大や細胞へのダメージの増加を招くこととトレードオフの関係にある。したがって、奥行方向の撮影枚数が限られている場合、対象物の詳細な観察や解析が困難になるという問題がある。

この問題に対して、拡散モデル [6] を用いた奥行方向の画像生成手法が提案されている [7]。拡散モデルは高品質な画像生成を可能とする一方で、高解像度の画像を生成するには計算コストが著しく増大するという課題がある。特に本研究の問題設定では、実時間での処理や大量のデータを扱う場合、莫大な計算資源が大きな制約となるため、計算量の削減が求められている。

本研究では、拡散モデルとニューラル場を組み合わせることで、共焦点画像の奥行補間を高速かつ高精度に実現す

る手法を提案する。具体的には、まず拡散モデルを用いて低解像度の画像を生成し、奥行方向に粗く補間する。低解像度の画像を扱うことで、拡散モデルが内部で繰り返し実行する各ステップの処理コストが低減し、ステップ数分の計算負荷を大幅に抑えられる。次に、拡散モデルで生成した低解像度の画像と元の共焦点画像を組み合わせた混合解像度データセットに対して、ニューラル場をフィッティングさせることで、画像平面と奥行の両方向において高解像度な共焦点画像を出力する。このアプローチにより、拡散モデル単体での画像生成と比較して計算時間を大幅に削減しつつ、高品質な共焦点画像を生成することが可能となる。

提案手法の有効性を検証するために、実験において従来の拡散モデルによる奥行補間との比較および評価を行った。その結果、提案手法は計算コストを大幅に削減しながら、生成した画像の品質において優れた性能を示した。

## 2. 関連研究

### 2.1 拡散モデル

拡散モデル [6] は生成モデルの一種であり、画像生成や音声生成などのタスクで広く利用されている。拡散モデルはガウシアンノイズを加えてデータを徐々に破壊する「拡散過程」と、その逆にノイズを除去してデータを復元する「逆拡散過程」の2つのプロセスを基盤としている。拡散過程では、入力データに対して段階的にランダムノイズを付加し、最終的には完全なノイズへと変換する。この拡散過程を数式で表すと以下ようになる。

<sup>1</sup> 奈良先端科学技術大学院大学  
Nara Institute of Science and Technology

<sup>2</sup> 金沢大学  
Kanazawa University

a) ohashi.tomoya.ou3@is.naist.jp

$$\mathbf{x}_t = \sqrt{1 - \beta_t} \mathbf{x}_{t-1} + \sqrt{\beta_t} \mathbf{z}_{t-1} \quad (1)$$

ここで、 $\mathbf{x}_t$  は時刻  $t$  におけるデータの状態、 $\beta_t$  は時刻  $t$  におけるノイズスケールパラメータ、 $\mathbf{z}_{t-1} \sim \mathcal{N}(0, \mathbf{I})$  は標準正規分布に従うノイズであり、 $\mathbf{I}$  は単位行列を意味する。拡散過程は正規分布を用いた確率的なものであり、逐次的に進行する。一方で、逆拡散過程では、拡散過程の最終ステップである完全なノイズ画像から元のデータを復元する。この逆拡散過程を数式で表すと以下ようになる。

$$p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_{\theta}(\mathbf{x}_t, t), \Sigma_{\theta}(\mathbf{x}_t, t)) \quad (2)$$

ここで、 $\mu_{\theta}(\mathbf{x}_t, t)$  はニューラルネットワークにより予測される平均、 $\Sigma_{\theta}(\mathbf{x}_t, t)$  は予測される分散である。逆拡散過程では、ノイズ除去の条件付き確率分布をニューラルネットワークを用いて学習する。この学習プロセスにより、モデルはノイズからデータを生成する能力を獲得する。

拡散モデルの関連研究として、潜在拡散モデル [12] が挙げられる。この研究では、従来の拡散モデルがピクセル空間を扱うのに対し、事前学習されたオートエンコーダの潜在空間で拡散モデルを適用することで、計算コストを削減しつつ高品質な画像生成を実現している。本研究に関連するものとしては、3次元の生物医学的ボリュームデータの高解像度化を2次元の高解像度顕微鏡データを用いて実現する手法 [7] も報告されている。

## 2.2 ニューラル場

ニューラル場とは、近年コンピュータビジョンやコンピュータグラフィックスの分野で大きな注目を集める技術であり、任意の座標を入力として、その座標に対応する値（色や密度などの物理量）をニューラルネットワークが出力する枠組みを指す [1], [3], [4], [10], [11], [13]。従来のボクセルのような離散的なデータ表現では、データの解像度を上げるとメモリ使用量が多項式的に増大し、情報を保存するために膨大なメモリが必要になる。しかしながら、ニューラル場を用いると、ニューラルネットワークのパラメータとして情報を圧縮的に格納できるため、少ないメモリで高精細な表現が可能になるという利点がある。

ニューラル場のうち代表的なものとしては、NeRF (Neural Radiance Fields) [10] が挙げられる。NeRF は、3次元座標と光線の方向（方位角と仰角）を多層パーセプトロンに入力し、対応する密度と色を出力する。学習には、異なる視点から撮影された複数の画像を用い、それらを再構成するようにパラメータを更新する。学習後は、元の学習データには含まれない任意の視点からの画像を生成することが可能である。

また、グリッド表現を持つニューラル場 [1], [4], [11], [13] として、K-Planes [3] が挙げられる。K-Planes は NeRF [10] のように多層パーセプトロンのみを用いる手法とは異な

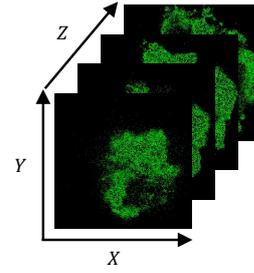


図 1 撮影された共焦点画像の例

り、グリッド上に直接特徴ベクトルを保持することで、高速な学習を実現する。加えて、K-Planes では任意の  $N$  次元グリッドを複数の 2 次元平面の組み合わせとして表現することで、メモリ使用量の削減を可能としている。

さらに、ニューラル場は本研究で対象とする共焦点画像における 3 次元情報の推定やノイズ軽減の研究 [9] にも応用されている。

## 3. 提案手法

### 3.1 共焦点画像

共焦点画像  $V \in \mathbb{R}^{X \times Y \times Z}$  は対象とする 3 次元空間を断層的に撮影したものであり、座標  $(x, y, z)$  における値は  $V(x, y, z) \in \mathbb{R}^3$  である。ここで、 $(x, y)$  は各画像のピクセル位置であり、 $z$  は撮影した奥行の位置である。例えば、図 1 の例は、立体組織であるオルガノイド [2] を奥行  $z$  方向に撮影した共焦点画像である。

### 3.2 手法の概要

本研究では、MSDSR [7] で生成した低解像度画像に対して K-Planes [3] をフィッティングさせることで、共焦点画像の高速な奥行補間を実現する手法を提案する。提案手法の流れは以下である。また、概要を図 2 に示す。

- (1) 解像度を 1/4 倍にリサイズした共焦点画像を用いて MSDSR の学習を行う。ここで、元の共焦点画像の解像度は  $256 \times 256$  であり、リサイズ後の解像度は  $64 \times 64$  である。
- (2) (1) で学習した MSDSR を用いて、入力となる元の共焦点画像 4 枚を補間する解像度  $64 \times 64$  の共焦点画像を生成する。これにより、低解像度の共焦点画像が 4 枚得られる。
- (3) 解像度  $256 \times 256$  の元画像 4 枚と、(2) で生成した解像度  $64 \times 64$  の低解像度画像 4 枚を組み合わせた計 8 枚の共焦点画像を用いて、K-Planes で画像を再構成するように学習する。
- (4) 学習後は任意の空間座標と奥行を入力することで、元の共焦点画像と同じ解像度で、もとのデータセットには存在しない奥行の画像を生成することができる。

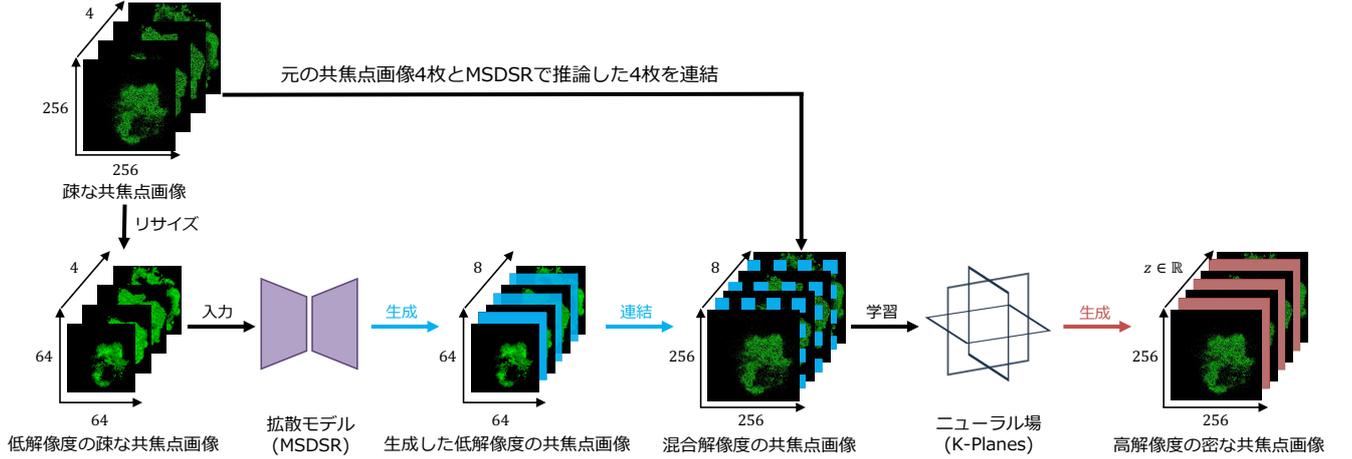


図 2 提案手法の概要図。手法は前半部分である (a) MSDSR による奥行方向の粗い補間と後半部分である (b) K-Planes による画像平面と奥行の両方向の高解像度化から構成される。

### 3.3 拡散モデルによる共焦点画像の学習

本節では、拡散モデルである MSDSR [7] を用いて、低解像度の共焦点画像を学習する方法について述べる。拡散モデルは 2 次元の事前学習済みモデル [6] を初期値として用いるため、3 次元である共焦点画像の奥行補間を直接行うことはできない。そのため、この問題を解決するために共焦点画像を  $XZ$  と  $YZ$  の 2 つの平面に分割して学習することで、奥行補間から行補間の問題に帰着させる。以下では、MSDSR を用いた共焦点画像の学習プロセスについて詳述する。

元の共焦点画像  $V \in \mathbb{R}^{X \times Y \times Z}$  に対して、MSDSR の学習に用いる低解像度の共焦点画像を  $\tilde{V} \in \mathbb{R}^{\tilde{X} \times \tilde{Y} \times Z}$  とする。ここで、 $Z$  は拡散モデルで補完後の奥行方向の解像度に相当する。次に、このような共焦点画像を  $N$  個集めたデータセット

$$D = \{\tilde{V}_1, \tilde{V}_2, \dots, \tilde{V}_N\} \quad \text{where} \quad \tilde{V}_i \in \mathbb{R}^{\tilde{X} \times \tilde{Y} \times Z} \quad (3)$$

を考える。拡散モデルは 2 次元の平面画像を扱う必要があるため、各共焦点画像  $\tilde{V}_i$  に対して、 $XZ$  平面と  $YZ$  平面を切り出す。  $x$  座標を固定して得られる  $YZ$  平面を

$$\tilde{V}_i^{YZ}(\tilde{x}) = \left\{ \tilde{V}_i(\tilde{x}, \tilde{y}, z) \mid \tilde{y} \in \{1, \dots, \tilde{Y}\}, z \in \{1, \dots, Z\} \right\} \in \mathbb{R}^{\tilde{Y} \times Z} \quad (4)$$

$y$  座標を固定して得られる  $XZ$  平面を

$$\tilde{V}_i^{XZ}(\tilde{y}) = \left\{ \tilde{V}_i(\tilde{x}, \tilde{y}, z) \mid \tilde{x} \in \{1, \dots, \tilde{X}\}, z \in \{1, \dots, Z\} \right\} \in \mathbb{R}^{\tilde{X} \times Z} \quad (5)$$

と定義する。これらの各平面 (4), (5) を集約して新たなデータセット  $D'$  を構成する。すなわち、 $i \in \{1, \dots, N\}$  と  $\tilde{x} \in \{1, \dots, \tilde{X}\}$  の全組み合わせからなる  $YZ$  平面群と、

$i \in \{1, \dots, N\}$  と  $\tilde{y} \in \{1, \dots, \tilde{Y}\}$  の全組み合わせからなる  $XZ$  平面群をあわせた集合として、

$$D' = \left\{ \tilde{V}_i^{YZ}(\tilde{x}) \mid i = 1, \dots, N, \tilde{x} = 1, \dots, \tilde{X} \right\} \cup \left\{ \tilde{V}_i^{XZ}(\tilde{y}) \mid i = 1, \dots, N, \tilde{y} = 1, \dots, \tilde{Y} \right\} \quad (6)$$

と定義する。以上より得られたデータセット  $D'$  を用いて MSDSR を学習する。学習時には、データセット  $D'$  からランダムに 1 つの平面  $v \in D'$  をサンプリングして MSDSR に入力し、拡散過程および逆拡散過程のプロセスを通じて再生成された平面に対して誤差関数を計算する。提案手法では、奥行方向を粗い解像度で補間する。学習時には、入力となる奥行きと補間対象である奥行きを区別するため、 $XZ$ ,  $YZ$  平面それぞれの  $Z$  行に対して 2 値マスク  $b$  を導入し条件として入力に与える。これにより、MSDSR が条件付き生成を行う際に、条件として与えられた行以外はノイズ除去を学習し、条件部分は元の行情報を保つように学習が進む。

以降では、まず誤差関数の定義を示し、その後、マスク  $b$  と関数  $c(\cdot)$  の具体的な内容を示す。学習時に最小化する誤差関数  $\mathcal{L}$  を以下のように定義する。

$$\mathcal{L} = \mathbb{E}_{v, b, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), t} \left[ \|b * \epsilon - \epsilon_\theta(c(v_t, v, b), t)\|^2 \right] \quad (7)$$

ここで、 $*$  は要素ごとの積を意味し、行方向へのブロードキャストを伴う。  $v$  は  $D'$  からサンプリングされた平面画像であり、 $v_t$  は拡散過程により時刻  $t$  にノイズが付与された画像である。また、 $\epsilon \sim \mathcal{N}(0, I)$  は平均 0、分散 1 のガウシアンノイズ、 $b$  は後述の行方向マスクであり、 $\epsilon_\theta(c(v_t, v, b), t)$  はパラメータ  $\theta$  を持つモデルにより予測されたノイズである。

本手法では、学習時の条件として、元の平面画像の一部の行を利用する。このとき、行数を制御するため、

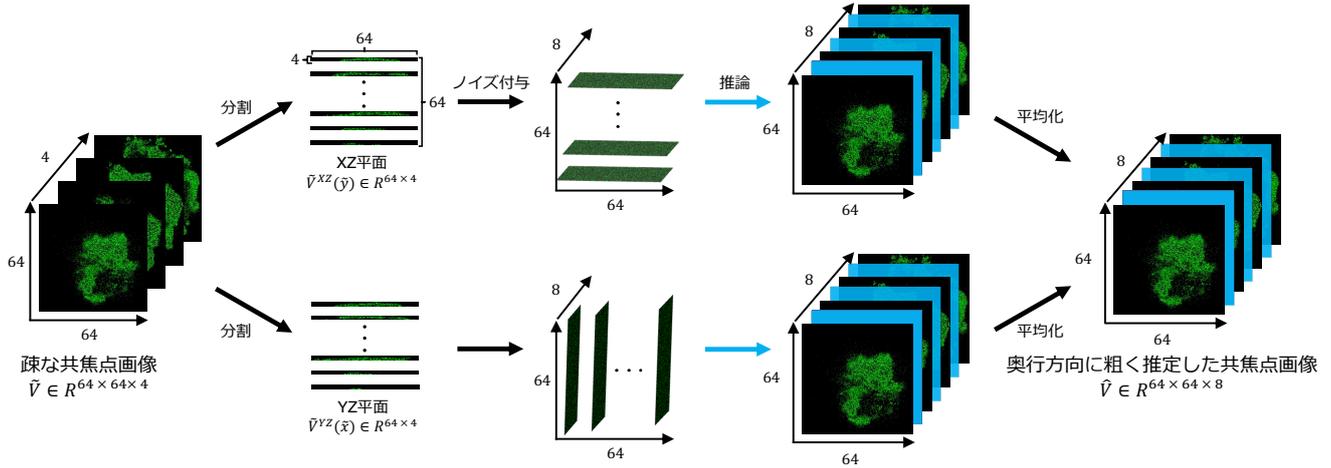


図 3 MSDSR による推論の全体図

$$0 < Z_{\min} < Z_{\max} < Z \quad (8)$$

という条件のもと、 $l \sim \text{Uniform}([Z_{\min}, Z_{\max}])$  により行数  $l$  をサンプリングする。さらに、 $\mathcal{S} \sim \text{Uniform}([1, Z], l)$  として、 $1 \leq i \leq Z$  のうち  $l$  個の行インデックスを無作為に抽出する。これにより、行方向の二値マスク  $b$  を以下のように定義する。

$$b = [\mathbf{1}_{\mathcal{S}}(1), \dots, \mathbf{1}_{\mathcal{S}}(Z)]^T \quad (9)$$

特性関数  $\mathbf{1}_{\mathcal{S}}(i)$  は以下のように定義される。

$$\mathbf{1}_{\mathcal{S}}(i) = \begin{cases} 1 & (i \in \mathcal{S}) \\ 0 & (i \notin \mathcal{S}) \end{cases} \quad (10)$$

すなわち、 $\mathcal{S}$  に含まれる行は 1、そうでない行は 0 としてマスクを作成する。これを用いて、関数  $c(\cdot)$  を

$$c(v_t, v, b) = b * v + (1 - b) * v_t \quad (11)$$

と定義する。直感的には、マスク  $b$  が 1 となる行は元の平面画像  $v$  を条件として残し、それ以外の行はノイズが付加された状態  $v_t$  の画像を用いることで、部分的な条件付き生成を実現する。

### 3.4 拡散モデルによる奥行方向の粗い推論

本節では、学習済み MSDSR を用いて、共焦点画像の奥行方向を粗く推論する過程について述べる。図 3 に MSDSR による推論の全体図を示す。

推論時には、まず入力される共焦点画像  $\tilde{V} \in \mathbb{R}^{\tilde{X} \times \tilde{Y} \times Z}$  を、式 (4) および式 (5) で定義されるとおり、XZ 平面群および YZ 平面群に分割する。これにより、MSDSR は各平面を独立に推論を行うことが可能となる。推論の際は、学習時と同様に行方向マスク  $b$  を用いて条件付き生成を行う。具体的には、逆拡散過程において観測されている層に対応する行はマスク  $b$  で 1 を与え、その他の行には 0

を割り当てることで、式 (9) および式 (11) に基づきマスクが 0 である行を MSDSR により推定する。なお、本研究では推論時は偶数行 ( $z = 2k, k \in \mathbb{N}$ ) を条件、奇数行 ( $z = 2k + 1, k \in \mathbb{N}$ ) を生成対象とする。

XZ 平面方向と YZ 平面方向の推定結果  $\hat{V}_{XZ}, \hat{V}_{YZ}$  は、いずれも最終的には  $\tilde{X} \times \tilde{Y} \times Z$  の 3 次元ボリュームとして得られる。

$$\hat{V}_{XZ} \in \mathbb{R}^{\tilde{X} \times \tilde{Y} \times Z}, \hat{V}_{YZ} \in \mathbb{R}^{\tilde{X} \times \tilde{Y} \times Z} \quad (12)$$

これらは同一の 3 次元空間を異なる平面分割で推定したものであるため、対応する画素同士を平均化して最終的な推定ボリューム

$$\hat{V} = \frac{\hat{V}_{XZ} + \hat{V}_{YZ}}{2} \in \mathbb{R}^{\tilde{X} \times \tilde{Y} \times Z} \quad (13)$$

を得る。図 4 は実際に生成した画像である。

### 3.5 ニューラル場による共焦点画像の学習

提案手法の後半部分では、MSDSR で推定した低解像度の画像と元画像を組み合わせた共焦点画像を再構成するように K-Planes [3] の学習を行う。

#### K-Planes

K-Planes で 3 次元空間を表現する場合は、互いに直交する 3 枚の 2 次元平面  $\mathbf{P}_{XY}, \mathbf{P}_{YZ}, \mathbf{P}_{ZX}$  を用いる。各平面上には  $M$  次元の特徴ベクトルを保持する。例えば、 $X, Y$  を各次元の解像度とすると、 $\mathbf{P}_{XY} \in \mathbb{R}^{X \times Y \times M}$  である。いま、ある 3 次元点  $\mathbf{q} \in \mathbb{R}^3$  を入力した場合、各平面  $c \in C$ 、 $C = \{XY, YZ, ZX\}$  上の特徴ベクトル  $f(\mathbf{q})_c \in \mathbb{R}^M$  を以下の式に従って取得する。

$$f(\mathbf{q})_c = \psi(\mathbf{P}_c, \pi_c(\mathbf{q})) \quad (14)$$

ここで、 $\pi_c$  は 3 次元点の 2 次元平面上への射影であり、 $\psi$  はバイリニア補間である。その後、アダマール積を用いて特徴ベクトル  $f(\mathbf{q}) \in \mathbb{R}^M$  を計算する。

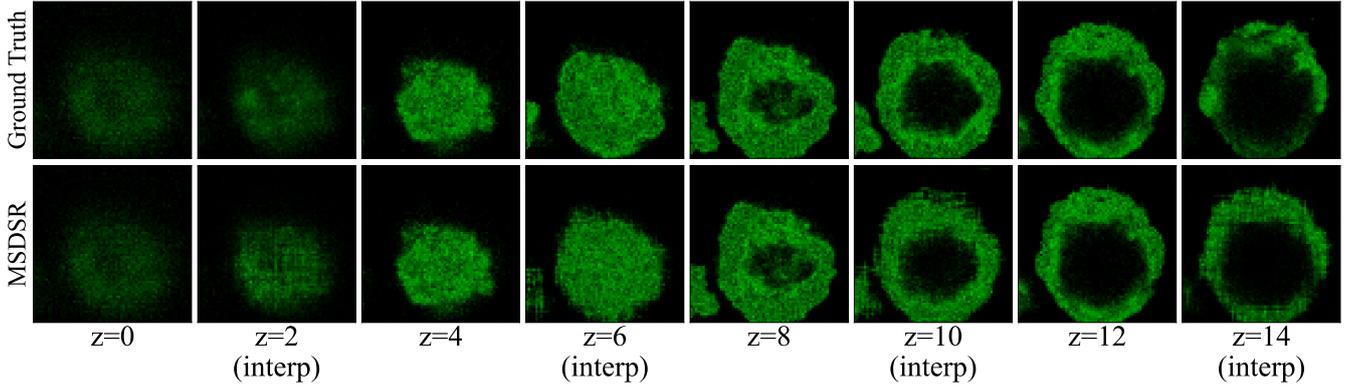


図4 MSDSRによる奥行方向の推定結果. 上段は実際に撮影された共焦点画像をMSDSRの出力サイズに合わせてリサイズした画像(正解画像), 下段はMSDSRで生成した画像である.  $z \in \{0, 4, 8, 12\}$  は条件として使用した画像であり,  $z \in \{2, 6, 10, 14\}$  はMSDSRで補間した画像である(“interp”と表記している).

$$f(\mathbf{q}) = \prod_{c \in C} f(\mathbf{q})_c \quad (15)$$

最後に, この特徴ベクトルを浅い多層パーセプトロン  $g$  に入力して値  $V^*(\mathbf{q}) \in \mathbb{R}$  を得る.

$$V^*(\mathbf{q}) = g(f(\mathbf{q})) \quad (16)$$

### K-Planes による共焦点画像の学習

学習データには, 元の共焦点画像と拡散モデルで生成した低解像度の画像を組み合わせた混合解像度データセット  $D''$  を用いる.  $D''$  は

$$D'' = \left\{ V(z) \mid z = 2k, k \in \mathbb{N} \right\} \cup \left\{ \hat{V}(z) \mid z = 2k+1, k \in \mathbb{N} \right\} \quad (17)$$

と定義する. ここで,  $V(z)$  は元の共焦点画像であり,  $\hat{V}(z)$  はMSDSRで推定した低解像度画像である. K-Planesを学習する際は, 低解像度画像  $\hat{V}(z)$  は元の共焦点画像  $V(z)$  の座標系にマッピングして学習する. 具体的には,  $\hat{V}(\tilde{x}, \tilde{y}, z)$  の座標  $(\tilde{x}, \tilde{y})$  に対してスケーリング係数  $s = R_{\text{high}}/R_{\text{low}}$  を乗算し, 元の座標系  $(x, y)$  に変換する. ここで,  $R_{\text{high}}$  は元の共焦点画像の解像度,  $R_{\text{low}}$  は低解像度画像の解像度を表す. つまり, 低解像度画像  $\hat{V}(z)$  の座標  $(\tilde{x}, \tilde{y}, z)$  における値はマッピング後には  $\hat{V}(\frac{x}{s}, \frac{y}{s}, z) \in \mathbb{R}^3$  になる.

元の共焦点画像の画像平面の解像度が  $X \times Y$ , 奥行方向の解像度が  $Z$  であるとき, 画像方向と奥行方向に対応するように3つの平面  $\mathbf{P}_{XY} \in \mathbb{R}^{X \times Y \times M}$ ,  $\mathbf{P}_{YZ} \in \mathbb{R}^{Y \times Z \times M}$ ,  $\mathbf{P}_{ZX} \in \mathbb{R}^{Z \times X \times M}$  を初期化する. 学習時は  $(x, y, z) \in \{1, \dots, X\} \times \{1, \dots, Y\} \times \{1, \dots, Z\}$  をランダムにサンプルし, K-Planesに入力して得られた出力に対して誤差関数を計算する.

### 誤差関数

K-Planesの学習には以下の再構成誤差を用いる.

$$\mathcal{L}_r = \frac{1}{XYZ} \sum_{x,y} \left( \sum_{k \in \mathbb{N}} (V(x, y, 2k) - V^*(x, y, 2k))^2 + \sum_{k \in \mathbb{N}} (\hat{V}(\frac{x}{s}, \frac{y}{s}, 2k+1) - V^*(x, y, 2k+1))^2 \right) \quad (18)$$

この再構成誤差  $\mathcal{L}_r$  を最小化するように, 各平面上の特徴ベクトルと多層パーセプトロンのパラメータを更新する. K-Planesでは再構成誤差に加え, 平面上の特徴に対する正則化を加える. 具体的には, 以下のように各平面上での total variation を計算する.

$$\begin{aligned} \mathcal{L}_{TV} = & \frac{1}{XY} \sum_{x,y} (\|\mathbf{P}_{XY}^{x,y} - \mathbf{P}_{XY}^{x-1,y}\|_2^2 + \|\mathbf{P}_{XY}^{x,y} - \mathbf{P}_{XY}^{x,y-1}\|_2^2) \\ & + \frac{1}{YZ} \sum_{y,z} (\|\mathbf{P}_{YZ}^{y,z} - \mathbf{P}_{YZ}^{y-1,z}\|_2^2 + \|\mathbf{P}_{YZ}^{y,z} - \mathbf{P}_{YZ}^{y,z-1}\|_2^2) \\ & + \frac{1}{ZX} \sum_{z,x} (\|\mathbf{P}_{ZX}^{z,x} - \mathbf{P}_{ZX}^{z,x-1}\|_2^2 + \|\mathbf{P}_{ZX}^{z,x} - \mathbf{P}_{ZX}^{z,x-1}\|_2^2) \end{aligned} \quad (19)$$

最終的な誤差関数はこれらの和であり以下となる.

$$\mathcal{L} = \mathcal{L}_r + \lambda \mathcal{L}_{TV} \quad (20)$$

ここで,  $\lambda$  はハイパーパラメータであり, 本研究では  $\lambda = 0.001$  とした.

### 3.6 ニューラル場による画像平面と奥行方向の高解像度化

学習時は対象とする元の共焦点画像とMSDSRで推論した低解像度の画像に存在する奥行  $z \in \{1, \dots, Z\}$  のみを入力する. 学習後は任意の値を入力することができるため, 元の共焦点画像の画像平面の解像度と奥行  $z \in \mathbb{R}$  を入力することで, もとのデータセットに存在しない奥行の画像を元の共焦点画像の解像度で生成することができる. これは

実際には、式 (14) に従って平面上の特徴空間で線形補間を行ったあと、式 (16) の多層パーセプトロンによる非線形変換で画像を生成することである。

## 4. 実験

提案手法に対して、MSDSR, K-Planes, 線形補間の3つの手法と比較することにより、提案手法による補間の有効性を検証する。

### 4.1 データセット

眼領域や中枢神経系の発生において重要なマスター遺伝子 Rax を蛍光標識したマウス胚性幹細胞から、網膜オルガノイドが形成される過程を共焦点顕微鏡で撮影した。実験では、培養開始から11の時期 (Day4, Day4.5, Day5, Day5.5, Day6, Day6.5, Day7, Day7.5, Day8, Day8.5, Day9) に撮影された共焦点画像を用いた。培養日数が少ない時期は分化の初期であり、比較的均一な形状である一方で、日付が経過するにつれて複雑な立体構造をもつようになる。データセット全体としては、この11の時期において、各時期につき96個のオルガノイドを対象とし、合計1056組から構成される。本研究では、これら1056組から学習データとして968組、評価用テストデータとして10組を抽出した。

各共焦点画像は画像サイズが  $512 \times 512$  であり、奥行方向には15枚の画像を撮影した。画像の1ピクセルは  $2.486 \mu\text{m}$ 、奥行方向の間隔は  $17.920 \mu\text{m}$  である。

今回撮影した共焦点画像には、被写体の背景にノイズが散在するため、前処理として、大津の閾値処理によるノイズ除去を行った。さらに、各共焦点画像について、被写体の輪郭の重心に基づいて  $256 \times 256$  にクロップする処理も施した。本実験では定量評価のために、 $z \in \{0, 4, 8, 12\}$  の4枚を入力とし、 $z \in \{1, 3, \dots, 11\}$  を評価に利用した。

### 4.2 実験結果

#### 共焦点画像の補間

図5に、共焦点画像の補間結果を示す。1行目は入力画像、2行目は正解画像 (Ground-truth)、3行目はMSDSR、4行目はK-Planes、5行目は線形補間、6行目は提案手法 (Ours) で生成した画像である。 $z \in \{1, 3, \dots, 11\}$  は学習に使用していないテストデータであり、各手法により補間した結果を示している。線形補間は、近傍の画像を単純に足し合わせるだけであるため、特に  $z = 5$  で示すように多重像のように見えるアーティファクトが生じやすい。これに対し、K-Planes では  $z = 7$  で示すように画像間で形状が大きく変化する領域において、正解画像とかけ離れた形状を推定してしまう場合がある。また、MSDSR では輝度が暗いボケた画像が生成されており、正解画像とは大きな乖離がある。これは、拡散モデルが極端に明るい輝度と極端に暗い輝度が混在する画像を学習する際、平均輝度が中

間値 (約0.5) に偏ってしまう影響で、推論時に極端な輝度を再現できないという問題 [5] に起因していると考えられる。Ours は、 $z = 3$  や  $z = 5$  では他の手法と同じように正解画像とは異なる形状が生成されているものの、MSDSR 単体と比較して高品質な画像の生成が可能である。

#### 従来手法との比較

提案手法 (Ours) の有効性を確認するため、従来手法と定量的な比較を行った。今回扱う共焦点画像は背景領域が広く、被写体が画像の一部しか占めない。そのため、このまま比較すると、画像の大部分を占める背景領域の影響により、局所的な形状の違いを評価しづらい。そこで、定量評価を行うにあたり、後処理として大津の閾値処理により背景と被写体を分離し、背景部分を無視して比較を行った。

表1と表2に定量的な比較結果を示す。補間結果の構造的な特徴を評価するため、評価指標としてRMSEとLPIPSを用いた。ここで、RMSEとLPIPSは値が小さいほど精度が高いことを意味し、各奥行インデックスで最も精度が良い値を太字、2番目に良い値を下線で示している。表1に示すRMSEの結果では、Oursは全ての奥行インデックスにおいてMSDSRを大きく上回り、他の手法と比較して同等以上の精度を示した。一方で、表2のLPIPSを用いた比較においても、Oursは全ての奥行インデックスにおいてMSDSRを上回る精度を達成した。しかしながら、OursはK-Planesや線形補間と比較して、大きな精度向上は見られなかった。以上の結果より、提案手法は従来手法と同等の精度を示し、全体的な奥行インデックスでは高品質な画像生成を実現できていることがわかる。

#### 計算時間

提案手法とMSDSR単体での計算時間を比較した。表3は、両手法における計算時間の比較結果である。今回、MSDSRの学習にはNVIDIA A100 (80GB) を4枚、推論にはNVIDIA RTX A6000 (48GB) を1枚用いた。提案手法の推論時間は、MSDSRの推論時間に加え、K-Planesの学習と推論も含めた時間である。提案手法の学習時間はMSDSR単体よりも約6.3倍速く、推論時間は約2.1倍速い。以上より、提案手法では計算時間の大幅な削減を実現した。

## 5. まとめ

本研究では、拡散モデルとニューラル場を用いた共焦点画像の高速な奥行補間を提案した。MSDSRで奥行方向に粗く推定した低解像度の画像に対して、K-Planesをフィッティングさせることで、奥行と画像平面の両方向に高解像度な共焦点画像を生成できる。実験では、MSDSR単体の手法と比較して、計算時間を大幅に削減しながら、高品質な共焦点画像の生成を実現した。

今後の課題としては、提案手法の精度向上と、より多様な共焦点画像データセットを用いた評価が挙げられる。特

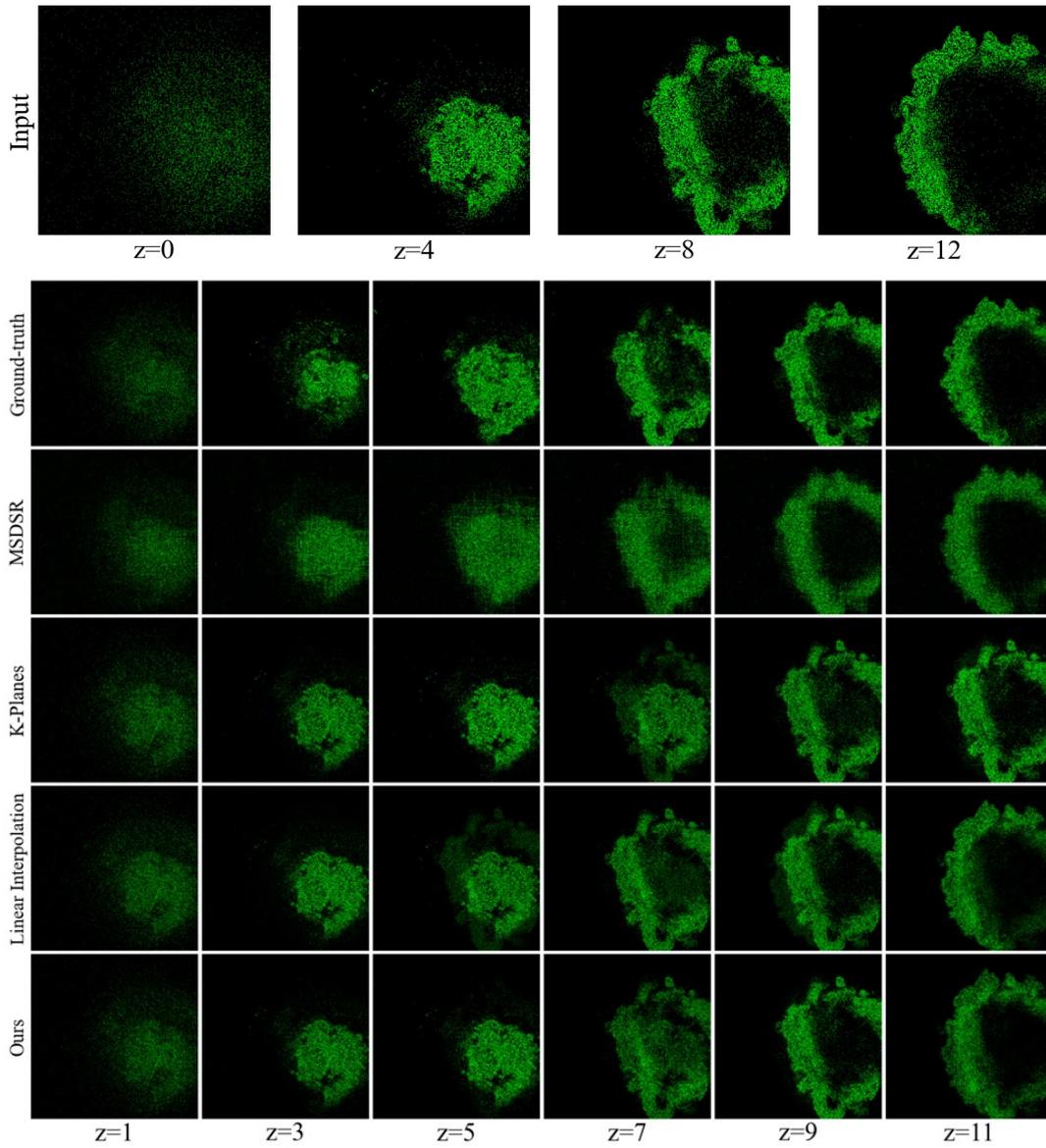


図 5 提案手法と従来手法による定性評価.  $z \in \{1, 3, \dots, 11\}$  は学習に使用しなかった画像であり、各手法により補間した画像である。

表 1 RMSE による評価

Method	$z = 1$	$z = 3$	$z = 5$	$z = 7$	$z = 9$	$z = 11$
MSDSR	0.221	0.254	0.281	0.281	0.272	0.257
K-Planes	<b>0.209</b>	0.221	<u>0.233</u>	0.269	0.223	0.239
Linear Interpolation	0.212	<b>0.218</b>	0.241	<b>0.237</b>	<u>0.219</u>	<b>0.225</b>
Ours	<u>0.211</u>	<b>0.218</b>	<b>0.227</b>	<u>0.259</u>	<b>0.212</b>	<u>0.237</u>

に、今回使用した実験データは明るさの分布が極端に偏っており、そのような画像の生成が困難であったと考える。この課題を克服するために、Kutsuna らの [8] で提案されているようなノイズスケジューリング手法を導入した拡散モデルを活用することで、輝度が極端に偏った画像であっても高品質な画像を生成できる可能性があると考えられる。

#### 謝辞

本研究は JST CREST JPMJCR23N3 (T.F.), JPMJCR1921

(S.O.), AMED 21bm0704065h0003 (S.O.), JSPS 科研費 JP22K17911 (Y.F.) の助成を受けたものである。

#### 参考文献

- [1] Chen, A., Xu, Z., Geiger, A., Yu, J. and Su, H.: TensorRF: Tensorial Radiance Fields, *ECCV* (2022).
- [2] Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., Sekiguchi, K., Adachi, T. and Sasai, Y.: Self-organizing optic-cup morphogenesis in three-dimensional culture, *Nature*, Vol. 472, No. 7341,

表 2 LPIPS による評価

Method	$z = 1$	$z = 3$	$z = 5$	$z = 7$	$z = 9$	$z = 11$
MSDSR	0.273	0.288	0.278	0.283	0.266	0.226
K-Planes	<b>0.244</b>	<u>0.232</u>	<b>0.225</b>	0.284	<u>0.194</u>	0.317
Linear Interpolation	0.252	<u>0.232</u>	0.238	<b>0.209</b>	0.201	<b>0.181</b>
Ours	<u>0.250</u>	<b>0.225</b>	<u>0.226</u>	<u>0.228</u>	<b>0.182</b>	<u>0.204</u>

表 3 計算時間の比較

	MSDSR	Ours
Training time	15638.1 sec	2464.7 sec
Inference time	649.7 sec	303.3 sec*

\* 303.3 sec = 14.7 sec (MSDSR) + 288.6 sec (K-Planes)

- pp. 51–56 (2011).
- [3] Fridovich-Keil, S., Meanti, G., Warburg, F. R., Recht, B. and Kanazawa, A.: K-Planes: Explicit Radiance Fields in Space, Time, and Appearance, *CVPR*, pp. 12479–12488 (2023).
- [4] Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B. and Kanazawa, A.: Plenoxels: Radiance Fields Without Neural Networks, *CVPR*, pp. 5501–5510 (2022).
- [5] Guttenberg, N.: Diffusion with offset noise, <https://www.crosslabs.org/blog/diffusion-with-offset-noise> (Accessed 2025/01/15).
- [6] Ho, J., Jain, A. and Abbeel, P.: Denoising Diffusion Probabilistic Models, *NeurIPS* (2020).
- [7] Jiang, C., Gedeon, A., Lyu, Y., Landgraf, E., Zhang, Y., Hou, X., Kondepudi, A., Chowdury, A., Lee, H. and Hollon, T.: Super-resolution of biomedical volumes with 2D supervision, *CVPRW* (2024).
- [8] Kutsuna, T.: Generalized Diffusion Model with Adjusted Offset Noise, *arXiv preprint arXiv:2412.03134* (2024).
- [9] Li, D.: Neural Rendering of Confocal Microscopy Images, *Proceedings of the 2022 5th International Conference on Digital Medicine and Image Processing* (2022).
- [10] 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, *ECCV* (2020).
- [11] Müller, T., Evans, A., Schied, C. and Keller, A.: Instant Neural Graphics Primitives with a Multiresolution Hash Encoding, *ACM TOG*, Vol. 41, No. 4, pp. 102:1–102:15 (2022).
- [12] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models, *CVPR* (2022).
- [13] Sun, C., Sun, M. and Chen, H.-T.: Direct Voxel Grid Optimization: Super-Fast Convergence for Radiance Fields Reconstruction, *CVPR*, pp. 5459–5469 (2022).