

フラクタル構造データを用いた三次元画像認識における深層学習モデルの事前学習

## Pre-training of Deep Neural Network for 3D Image Recognition using Fractal Structure Data

原田 航平<sup>†</sup> 野口 渉<sup>‡</sup> 田村 康将<sup>§</sup> 山本 雅人<sup>§\*</sup>  
Kohei Harada Wataru Noguchi Yasumasa Tamura Masahito Yamamoto

### 1. はじめに

三次元画像認識とは、三次元空間における物体について内部構造まで考慮して解析を行うタスクである。現実世界を構成する三次元構造を観察する際、表層のみでは情報が完結せず、なおかつ物理的に解剖して内部構造を観察することが難しい構造物がいくつか存在する。そうしたものの例として生体組織や工業部品などが例に挙げられる。そのような構造物に対して、CT や MRI といった非破壊的なイメージング技術を用いることで、奥行き情報を含む三次元の画像を構成することができる。こうした三次元画像は、内部構造の解析や診断、製品検査などにおいて極めて重要な役割を果たしている。その一方で、それらを三次元画像として機械学習へ応用することは、自然画像を中心とする二次元画像の応用ほどは進んでいない。その理由として、特殊な機器を用いることでしか撮影できない点や、倫理的制約から一般に公開されることが少ない点などによってデータの絶対数が少ないということが挙げられる。

本研究では、数式から生成される三次元フラクタル構造を活用することでこうした課題への対処可能性を検討する。フラクタル構造は簡単な数式から生成可能で、公平性や倫理的制約のような大規模データセットの抱える問題を解決できることから、膨大な事前学習が必要な二次元画像分類等のタスクにおける事前学習データとして活用されており、その有効性が確認されている[1, 2]。本研究では、これら先行研究の手法を三次元に拡張し、三次元のフラクタル構造をボクセルとして表現したデータセットで事前学習を行うことにより、データ収集コストの高い三次元実画像に対する汎用的な特徴抽出能力を獲得することを目指した。実験では、三次元画像分類タスクにおける三次元フラクタル構造の事前学習の有効性を検討している。事前学習を行ったモデルと行っていないモデルという 2 つのモデルに対して、それぞれ学習に用いるデータ数を変化させた場合における test データに対する精度の比較を行った。

### 2. 三次元フラクタル画像の生成と学習

本節では、フラクタル構造の生成に密接に関わる IFS と、

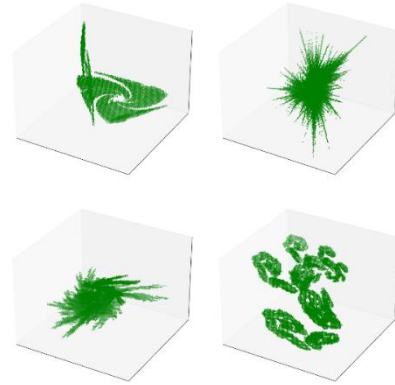


図 1 IFS によって生成されたフラクタル構造

実験に使用するデータセットの作成方法の詳細について述べる。

#### 2.1 Iterated Function System

三次元フラクタル構造のデータセットは IFS (Iterated Function System) を用いて生成した。IFS はフラクタル構造を生成するための数学的な枠組みの一つで、初期値となるある一点に対して、アフィン変換を適用する処理を繰り返すことにより、点の集合からなる自己相似性をもつ形状を生成する手法である。なお、適用するアフィン変換は、事前に定義された大きさ  $M$  の集合から繰り返しのたびに確率的に選択される。三次元空間上のアフィン変換は式 1 の通りである。

$$x_{n+1} = \begin{bmatrix} a_m & b_m & c_m \\ d_m & e_m & f_m \\ g_m & h_m & i_m \end{bmatrix} x_n + \begin{bmatrix} j_m \\ k_m \\ l_m \end{bmatrix} \quad (1)$$

IFS の  $n$  回目の試行では、まず  $M$  個のアフィン変換の集合から要素の 1 つを選択する。点  $x_{n+1}$  は、選択されたアフィン変換により、線形変換行列と  $x_n$  の積に平行移動ベクトルを加えることによって計算される。したがって IFS は行列とベクトルからなる  $12 \times M$  個のパラメータで定義される。実際に IFS を用いて生成したフラクタル構造を図 1 に示す。

#### 2.2 IFS を用いたデータセットの構成

本研究におけるデータセットは IFS により生成した三次元フラクタル構造からなる。IFS によって生成されるフラクタル構造はアフィン変換のパラメータに依存しており、異なるパラメータにより生成された結果は、それぞれ別カテゴリのフラクタル構造として扱うことができる。一方で、変化するパラメータが微小量である場合、収束するフラクタル構造の差異は小さくなることが知られており、その性質を利用することでカテゴリ内のインスタンス数を拡張す

<sup>†</sup> 北海道大学大学院情報科学院  
Graduate School of Information Science and  
Technology, Hokkaido University

<sup>‡</sup> 北海道大学数理・データサイエンス教育研究センター  
Education and Research Center for Mathematical  
and Data Science, Hokkaido University

<sup>§</sup> 北海道大学大学院情報科学研究院  
Faculty of Information Science and Technology,  
Hokkaido University

\* 北海道大学人間知・脳・AI 研究教育センター  
Center for Human Nature, Artificial Intelligence, and  
Neuroscience, Hokkaido University

ることができる[1]. 今回の三次元フラクタル構造の生成では, 各アフィン変換の 12 種のパラメータについて係数をかけることによる拡張を行っている. 係数のパターンは 12 通り用意し, それらを 12 種のパラメータのどれか 1 つに適用する. 結果として, 合計  $12 \times 12$  個のインスタンス拡張が行われ, 係数を一切かけないオリジナルのものも含めて 1 つのカテゴリあたり 145 個のインスタンスが定義される. IFS による 1 回の生成ごとに 100,000,00 点の三次元座標を取得しており, それらを  $[0, 127] \cap \mathbf{Z}$  に正規化を行うことで  $128 \times 128 \times 128$  の大きさのボクセルで表現される三次元画像を定義している.

本実験では上記のように定義されるフラクタル構造の三次元画像を 200 カテゴリ生成した. それらを 100 カテゴリずつ 2 つのデータセットに分割して, 一方を事前学習用のデータセットとして定義し, 他方を実際の三次元画像分類タスクにおける分類対象のデータに見立てた検証用のデータセットと定義した. 検証用のデータセットは 8:2 の割合で train データと test データに分割しており, モデルの精度検証においては学習に使用しない test データを用いて行った.

### 3. 実験

三次元フラクタル構造データの事前学習の有効性を確認するため, 次の 2 つのモデルの精度の検証を行った. 1 つは検証用データセットの train データのみを用いて学習を行ったスクラッチモデルである. もう 1 つは事前学習用データセットで十分学習したのち, 検証用データセットの train データを用いてファインチューニングを行ったモデルである. これら 2 つのモデルについて, それぞれ使用する train データの数を一定の割合に制限した場合における精度を検証することによって, データ数が少ない場合における事前学習の有効性を確認する.

また, 本実験では三次元画像分類タスクを行うためのアーキテクチャとして ViT を使用した[3]. 通常の ViT は二次元画像を入力とするため, パッチ分割の際に一辺の長さをパッチサイズとする正方形の大きさ事に埋め込みを行うが, それを三次元画像入力に拡張するために, 深さ方向の次元を考慮して一辺の長さをパッチサイズとする立方体の大きさごとに埋め込みを行った.

#### 3.1 モデルの学習

2 つのモデルについてはいずれも Cross Entropy Loss を損失関数とし, 最適化アルゴリズムに Adam 使用して損失が収束するまで訓練された. 使用する検証用データセットの train データの割合について, 25%と 50%と 75%と 100%の 4 通りに変化させた場合の学習をそれぞれ行っており, 計 8 通りのモデルを作成した.

なお, 事前学習を行ったモデルについて train データでファインチューニングする際には, クラス分類を行う全結合層である MLP Head 以外のパラメータは凍結して学習を行った.

#### 3.2 モデルの精度評価

使用した検証用データセットの train データの割合を横軸, test データに対する精度を縦軸にとり, モデルごとにプロ

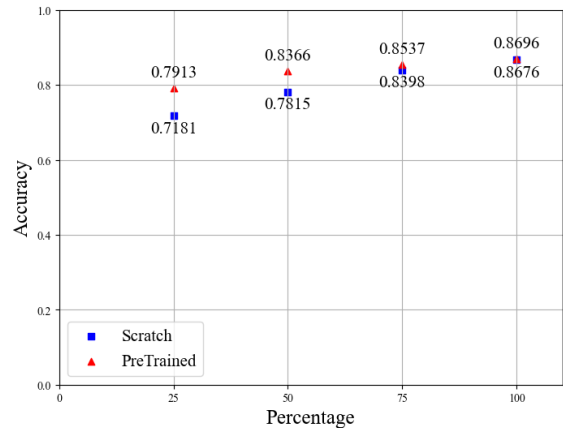


図 2 データ数ごとの各モデルの精度

ットしたのが図 2 である. これを見るといずれのパーセンテージにおいても事前学習を行ったモデルの方がスクラッチモデルよりも優れていることがわかる. また, データ数が少なくなるほど精度の差が顕著に現れる傾向にあり, 今回のタスクでは事前学習ありで train データを 50%使用したモデルと, train データを 75%使用したスクラッチモデルがほぼ同程度の精度を示す結果となった. 事前学習を行ったモデルのファインチューニングの際には MLP Head 以外のパラメータを凍結していることから, 事前学習により汎用性の高い特徴抽出能力が獲得できたことが推察される. 一方で, 使用する train データの割合が 75%と 100%の場合においては 2 つのモデルの差異は小さくなっており, スクラッチモデルであってもデータ量が十分であれば適切な特徴抽出能力を獲得し得ることがうかがえる.

このような結果はデータ数の少ない三次元実画像への応用を考えた場合に, 三次元フラクタル構造を用いた事前学習が効果的なアプローチであることを示唆し, その学習の汎用性からさまざまな分野への応用が期待される.

### 4. おわりに

本研究では, 数式からフラクタル構造をもつ三次元画像データセットを作成し, それらを用いて事前学習を行ったモデルの評価を行った. 事前学習を行ったモデルは, スクラッチモデルに比べてより少ない学習データ数で高い精度を示す結果となり, 三次元画像においてもフラクタル構造データによる事前学習の有効性が確認できた.

一方で, 今回有効性が確かめられたのは生成した三次元画像の範囲のみであり, 分布の異なるデータへの性能は未検証であるため, 今後は実画像を用いたタスクへの適用を目指す.

#### 参考文献

- [1] Kataoka, Hirokatsu, et al. "Pre-training without natural images." Proceedings of the Asian Conference on Computer Vision. (2020).
- [2] Yamada, Ryosuke, et al. "Mv-fractaldb: Formula-driven supervised learning for multi-view image recognition." 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, (2021).
- [3] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." International Conference on Learning Representations (2021).