

TSUBAME 共同利用 令和7年度 学術利用 成果報告書

深層学習に基づく小惑星上のボルダー検出技術の研究開発
Development of a Deep Learning-Based Technology for Boulder Detection on Asteroids

本田親寿¹, 許 峻豪², 富 宣超², 神山徹³, 吉川一朗²
Chikatoshi Honda¹, Hur Junho², Fu Xuanchao², Toru Kouyama³, and Ichiro Yoshikawa²

¹ 公立大学法人会津大学, ² 東京大学, ³ 産業技術総合研究所

¹The University of Aizu, ²The University of Tokyo, ³National Institute of Advanced Industrial Science and Technology

邦文抄録(300 字程度)

未知の小惑星におけるボルダー検出において、従来の深層学習モデルの汎用性の低さが大きな課題となっている。本研究では、膨大なデータセットで事前学習された基盤モデル「Segment Anything Model (SAM)」をファインチューニングすることで、この問題の解決を図った。検証の結果、提案モデル(FT-SAM)は従来の CNN ベースのモデルと比較して、未学習の天体データセットにおいて再現率を 30 ポイント以上向上させ、検出数は約 4 倍に達した。また、算出された累積ボルダーサイズ分布は先行研究の知見と高い相関を示した。本手法は基盤モデルの高い特徴抽出能力を活用することで、将来の小惑星探査における自動解析システムの構築を大幅に加速させるものである。

英文抄録(100 words 程度)

The poor generalization of deep learning models for boulder detection on unseen asteroids remains a key challenge for exploration missions. We address this by fine-tuning the "Segment Anything Model" (SAM), a foundation model pre-trained on a massive dataset. Compared to conventional CNN-based architectures, our fine-tuned SAM (FT-SAM) demonstrated superior robustness, improving the recall rate by over 30 percentage points and detecting four times more boulders on untrained asteroid datasets. The resulting boulder size-frequency distributions showed a high correlation with previous studies. This approach leverages pre-trained features to significantly enhance cross-domain performance, accelerating automated analysis for future asteroid missions.

Keywords: Fine Tuning, Deep Learning, Foundation Model, Transformer, MaskRCNN.

背景と目的

宇宙技術の革新により宇宙から地球に送信されるデータ量が増えている。「はやぶさ2」や「OSIRIS-REx」、DART ミッションなどの小惑星探査ミッションにより、小惑星表面の高解像度画像が多数取得され、科学的議論を可能とさせる豊富なデータセットが整備されつつある。これにより、小惑星表面に存在するボルダーのサイズ、形状、方位などの統計的特徴を精密に解析し、天体の形成過程や進化史を比較研究することが可能になる(cf. Michikami et al., 2021)。たとえば、Ryugu と Bennu は一見するとアルベド、スペクトル、密度、さらにはリターンサンプルの CI 的な特徴を共有してもつことが見られるが、ボルダーの形状分布や方位の特性には違いが認められている。これらの違いは、それぞれの天体が異なる母天体由来であることや、形成時の物質分布の不均質性、あるいは進化経路の違いを示唆している可能性がある。が、しかし、そのデータを使

った理学的成果の創出に目視に頼る過程が存在する。例えば、小惑星サンプル採集計画でタッチダウン場所探索や小惑星の形成過程を解明するために必要な表層環境の調査(Michikami et al., 2019)やクレーター年代調査における計数と採寸測等では目視解析を第一としていた(Hirata et al., 2019)。現時点まで小惑星リュウグウの表層画像の僅か4%までか解析できなかったように、人間が大量のデータを処理することは時間的・経済的に限界がある。その限界を乗り越えるために深層学習を導入することで問題の解決をはかる研究も進んでいるが、従来の深層学習モデルでは訓練した天体でしか使えない、正確度が低い等の問題があることから実用レベルで深層学習を惑星科学に取り入れることは難しい。そこで、私は観測対象が変わっても頑強で、人間と同等以上の画像認識能力が報告されている基盤モデル技術(Bommasani, et al., 2021)に着目し、画像内物体を認識・採寸する機械学習モデルの構築

が必要であると考えた。基盤モデルとは、近年利用が広がっている ChatGPT のようにデータのパターンや関係を学習し、新しいコンテンツを生成できる「生成 AI」の核となる技術であり、従来の深層学習ではできなかったデータ全体の文脈を学習することで性能を高めた Transformer 構造(画像の学習では Vision Transformer, ViT, Dosovitskiy et al., 2021)を多数組み込み、数億といった膨大なデータセットで事前に訓練された機械学習モデルのことを指す。基盤モデルは数枚のサンプルを少量学習(Fine-Tuning)することで性能を維持できながら異なる種別の解析を実施できる汎用性と、人間が分光・高度差等を観測対象を理解するために同時に使うように、複数の異なる情報を同時に組み合わせてモデルの入力として処理できる点に強みがある。

以上の観点から、本研究では一般写真において高い汎用性能が確認されている Meta 社が開発した基盤モデル「Segment Anything(SAM)」(Kirillov et al., 2023)の惑星探査データへの応用可能性を検討し、従来の CNN や Transformer が組み込まれた CNN モデルと比較することで性能を検証し、学習した天体と異なる天体に適用可能な汎用的な小惑星ボルダー検出モデル構築のためのパラメータを調査した。

概要

本研究では基盤モデルが持つ汎用的な特徴抽出能力を小惑星という未知ドメインに適応させることで、撮影条件や天体に左右されないモデルの実現を目的として多様かつ大規模な地球画像データセットで事前学習された基盤モデルを、少量の小惑星画像を用いてボルダー検出タスクに適するようにファインチューニング(少量学習)を行った。基盤モデルは既に 1100 万枚の画像で事前学習をしているため、強力な Image Encoder を凍結したまま、マスク生成を担う Mask Decoder のみをファインチューニングを行い、小惑星のボルダーの検出に特化したモデルの構築を試みた。

性能評価には Recall、IoU (Intersection over Union)、Overall Accuracy (OA) を用いて行い、小さなボルダーを見やすくするための画像のリスケールを行った。従来の CNN ベースのモデルと比べて学習した天体とは異なる天体に適用したところ、Recall Rate では 30 ポイ

ント以上上回る結果を示し、検出したボルダーの数は約4倍に至った。

本研究の結果は、従来の CNN ベースのモデルよりも基盤モデルを用いた解析が小惑星表面の画像解析、特に小さなボルダーの検出において高精度かつ頑固であると同時に、様々な天体にすぐ適用可能な汎用的なモデルを実装できることを示唆した。また、基盤モデルを Fine-tuning することで、複雑な地形や陰影が存在する小惑星表面においても基盤モデルが有する強い画像特徴抽出能力を活かし、少ない Ground Truth を用意するだけで強力なモデルを実装できることを確認した。

結果および考察

本研究では、深層学習モデルの学習や基盤モデルを少量学習するために先行研究で用いてリュウグウデータセットと学習したモデルの般化性能の評価するためにベヌーのデータセットを構築した。

○使用データと前処理

JAXA の小惑星探査機「はやぶさ2」に搭載された光学航法カメラ(ONC-T)が近接撮影した可視光域の画像のうち、228 枚を使用した。宇宙空間が写らないように小惑星表面までの距離が 5km 以下で撮影した画像に限定し、鮮明度向上のために Sigma Contrast によるコントラスト強調と、Wiener Deconvolution を用いたぼやけ除去の前処理を適用した。

NASA 小惑星探査機 OSIRIS-Rex に搭載された MapCam および SamCam で撮影された可視光域の画像から 高度 3km 以下で撮影された画像を制限した。また、図1のように緯度、経度、太陽入射角、高度、解像度が特定の範囲に偏らないよう、多様な条件下で撮影された画像を均等に選別し、データセットの多様性を確保した。

先行研究に基づき、検出精度を向上させるため、それぞれのデータセットに対して元画像(1024×1024)を4分割し、512×512 ピクセルのデータセットとして構築した。

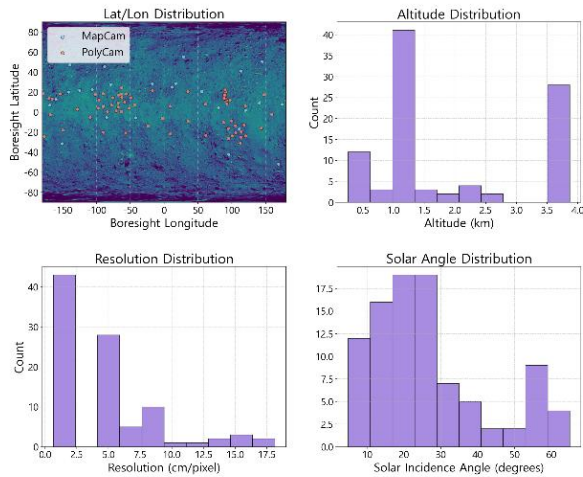


図 1 地理的な位置、高度、解像度、そして画像の太陽角度の分布を示す、多種多様な観測条件の要約です。

○モデル構造

本実験では、Simple Mask R-CNN フレームワークを利用した。本研究では、CNN ベースの ResNet-50 を Transformer ベースの SOD-Former に置き換えることで、微小なボルダーの検出に最適化されたバックボーン構造を構築している。低解像度の画像をダウンサンプリングした後、8 つの Residual Hybrid Attention Group (RHAG) モジュールを適用する手法を実装することで、より複雑な特徴の抽出を可能にするアーキテクチャとなっている(図 2 参照)。

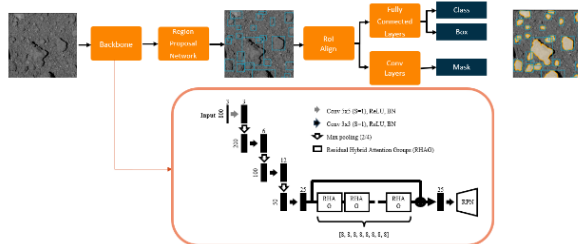


図 2 SOD-Former モデルの概要図。MaskR-CNN のバックボーンを SOD-Former に置き換えている

基盤モデルとして Meta 社の SAM (Segment Anything Model) を採用した。SAM ファインチューニングにおいては、同モデルが 1,100 万枚の画像で事前学習された強力な Image Encoder を保持しているため、マスク生成を担う Mask Decoder のみを学習対象とした。モデルの入力は、1 辺あたり 64 個のポイントと、各ポイントがボルダーか地表かを識別する情報で構成される。推論さ

れたボルダー情報にシグモイド関数を適用した後、誤差逆伝播を行った。

学習時、モデルには正解マスクと各グリッド点のラベル(前景/背景)を入力とし、その出力を Sigmoid 関数で 0 から 1 の確率マップに変換する。この確率マップを基に、クラス不均衡問題に有効な Focal Loss とマスク形状の精度を高める Dice Loss を組み合わせた損失関数を設計し、小さなボルダーも学習できるようにモデルを最適化した。このプロセスを図 3 に示す。

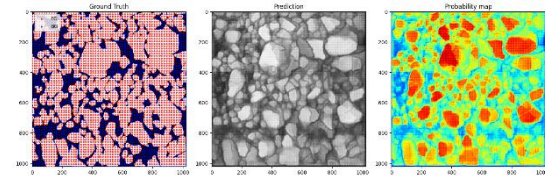


図 3 SAM のファインチューニングにおける、可視化されたモデル入力(左)、出力(中央)、および損失計算のために出力に適用されたシグモイド関数(右)

○評価指標と実験結果

モデルの性能評価には 3 つの指標を用いた。各評価指標の詳細は以下の通りである。なお、再現率 (Recall) における「正解 (Correct)」の判定基準は、IoU 値が 0.5 を上回る場合と定義した。

$$\text{検出率} = \frac{\text{GT に含まれるボルダーのうち正しく検出したボルダー数}}{\text{全てのボルダーの数}} \times 100$$

$$\text{IoU} = \frac{\text{GT とモデルのボルダーピクセルが一致する総数}}{\text{全てのピクセル数}} \times 100$$

$$\text{OA} = \frac{\text{バックグラウンドを含めて GT と一致するピクセル数}}{\text{全てのピクセル数}} \times 100$$

表 1 に各モデルの性能を示す。SOD-Former は未知の環境への適応に失敗し検出率が低迷した一方で、基盤モデル (Foundation models) は総じてボルダーを正確に捕捉する傾向が見られた。また、図 4 の可視化結果が示す通り、ファインチューニング前の SAM は単一の大型ボルダーを 2 つの独立した物体として誤認、あるいは検出に失敗する事例が見られたが、ファインチューニング後のモデル (FT-SAM) ではこれらの課題がいずれも解消されていることが確認された。

表 1 Benu データセットに関しては、リュウグウのデータセットのみを用いて学習し、ベンヌのデータセットで評価を実施した。赤字は各データセットにおける単一評価での最高性能を示す。

Model	Recall (%)	IoU (%)	OA (%)	# boulder
Ryugu Dataset				
SOD-Former	59.55	68.48	84.32	127
SAM	57.05	55.36	85.30	202
FT-SAM	68.25	62.01	87.25	372
Benu Dataset(Trained with only Ryugu dataset)				
SOD-Former	22.04	44.89	61.45	118
SAM	56.59	66.68	78.14	413
FT-SAM	75.43	81.60	87.55	781

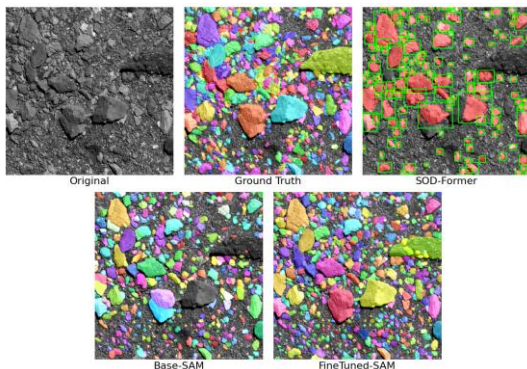


図 4 基盤モデルと Fine-tuning を行った基盤モデル、SODFormer を用いた推論画像の例

基盤モデルのファインチューニングにおける最適パラメータの検討も行った(図 5 参照)。ベンヌのデータを用いてベンヌを検出する性能は漸増するものの、同モデルをリュウグウのデータに適用した場合、学習の進行に伴い再現率が顕著な低下傾向を示すことが判明した。

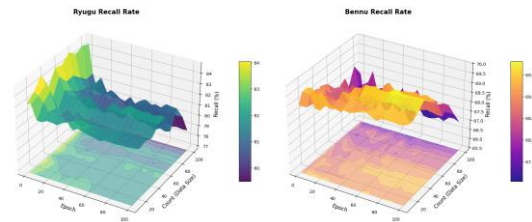
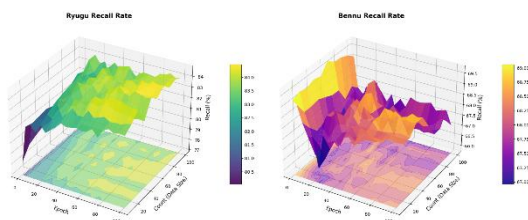


図 5 最適なファインチューニングパラメータ決定のためのエポック数および画像数変化に伴う評価。上段はリュウグウで学習し、リュウグウ・ベンヌ双方で検証した結果、下段はベンヌで学習し、双方で検証した結果を示す。

モデルの安定性を検証するため、ファインチューニング済み SAM を用いて累積ボルダーサイズ分布を算出した。その結果、図 6 に示す通り先行研究の知見と極めて高い相関が得られた。さらに、単位面積あたりのボルダー数が観測高度に応じて層状の分布(階層化)を示すことも確認された。

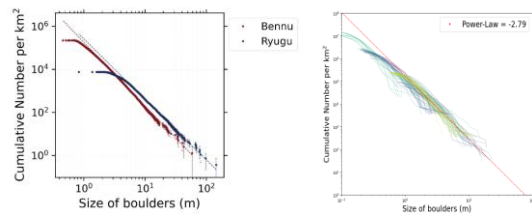


図 6 清水ら(2025)によるベンヌおよびリュウグウの全球累積サイズ頻度分布(左)と、ファインチューニング済み SAM モデルによる算出結果(右)

まとめ、今後の課題

本研究の結果、FT-SAM は少量の学習データを準備するだけで、他のモデルを上回る性能を示した。この結果は、基盤モデルの強力な Image Encoder が持っている特徴抽出能力が、少量学習を通じてボルダーという特定の対象物の検出に効率的に適応したと考えられる。つまり一つの小惑星データで学習したモデルを、Encoder 部分は再学習なしに他の小惑星にも適用できる可能性を示唆した。今後の課題として、大きいボルダーの検出率を向上させる必要があると同時に、異なる特性を持つ多様な天体に対しても同様の性能を維持できるかの検証が必要である。