

ARC-Space シンポジウム(12/19/2025)

基盤モデルのファインチューニングによる小惑星ボルダー 検出精度改善についての検証

Hur Junho^{1,2}, Toru Koyama², Xuanchao Fu^{1,2}, Ichiro Yoshikawa¹

1. Tokyo University

2. AIST

1.1 背景 – 小惑星探査機はやぶさ2、OSIRIS-REx

■ はやぶさ2

目標：小惑星リュウグウのサンプルリターン

カメラ装備：

- ・7つの分光カメラONC-Tを利用
- ・ $0.0006^\circ/\text{pix}$, 10cm/pix(高度1km)



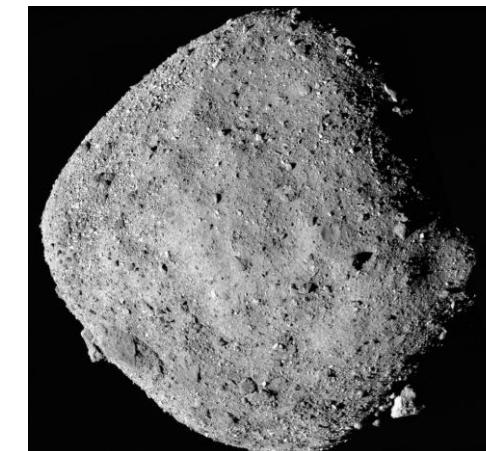
リュウグウの全休観測

■ OSIRIS-REx

目標：小惑星ベヌーのサンプルリターン

カメラ装備：

- ・3個のカメラ(MapCam, PolyCam, SamCam)を利用
- ・ $0.00075^\circ/\text{pix}$, 5cm/pix(高度1km, PolyCam)

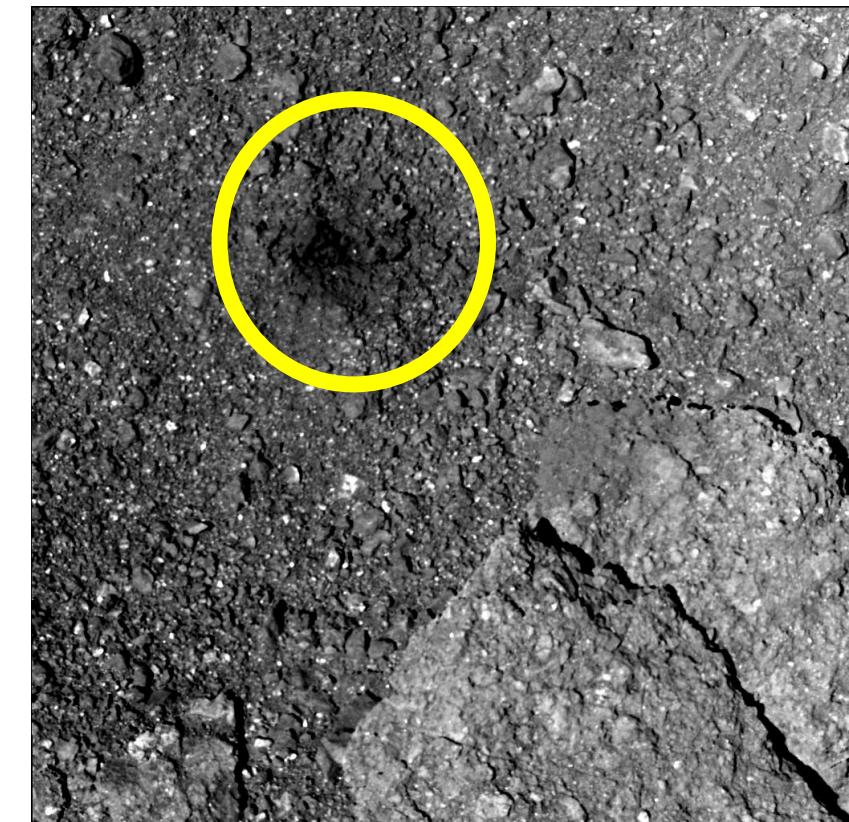


ベヌーの全休観測

1.2 背景 – 小惑星の形成過程と小惑星探査の意義

■ 小惑星探査の重要性

- 小惑星の形成と進化過程を理解できる
- 小惑星のボルダーは初期太陽系の物質を保存
→太陽系・惑星の歴史の解明に繋がる



■ ボルダー研究の重要性

- 探査機の着陸地点選定
- 小惑星表面の物理的特性から進化過程推測
- 衝突歴史の証拠収集

50cm

1-3 背景 - ボルダーに関する先行研究 (地域特性の定量化に用いられる指標)

数密度 (1km²あたりのボルダーの数)

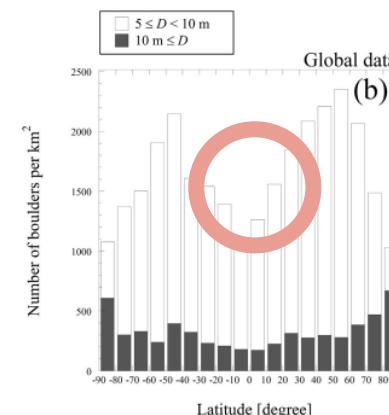
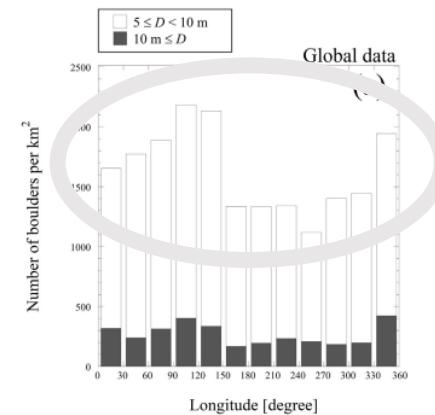
■ 緯度、経度依存性の確認

経度160°~290°Eの密度

→集積の際の偏り？

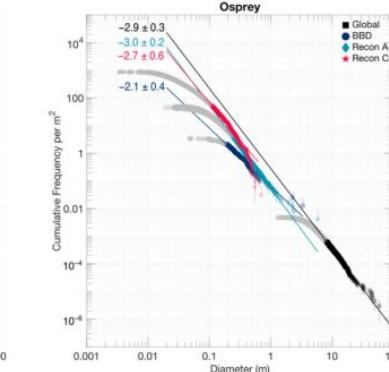
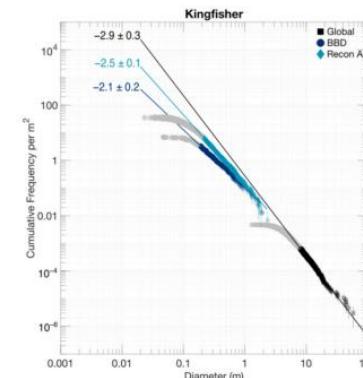
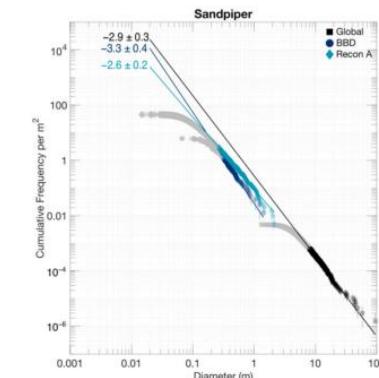
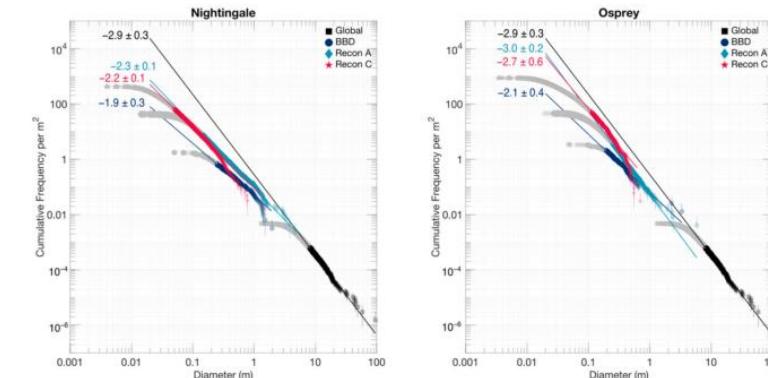
赤道の数密度が低さ

→細かい粒子で埋もれて
いる？



累積サイズ頻度分布

■ グラフの傾きが急であるほど、ボルダー破片化が進んでいると推定可能



1-4 背景 – ボルダー解析における課題

■ 理学的背景・課題

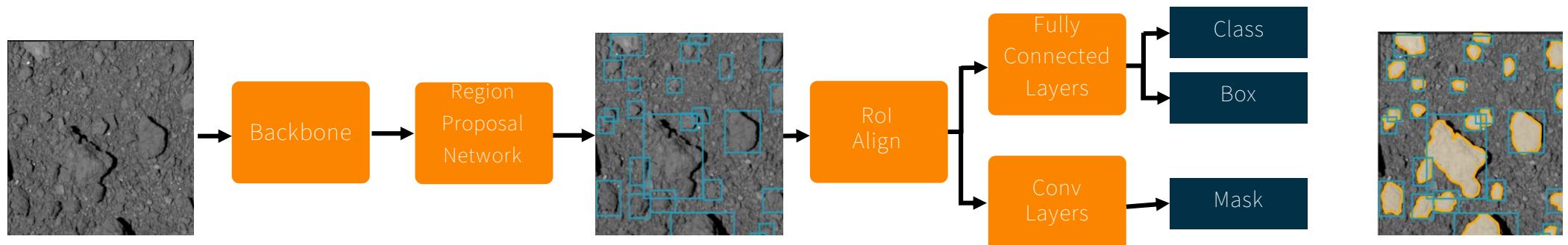
- ボルダーの分布の調査から、ボルダーが経験した地質学的プロセスが明らかになる
- リュウグウ全球 2.7km^2 上の 0.1m 以上のボルダーはおよそ 10^7 個
→ すべてのボルダーを調査するには膨大な時間がかかり、全球規模の解析はできていない
- 全球規模の解析を実現するために、機械学習を用いてボルダーを検出する手法を開発する

■ 機械学習を用いたボルダー検出の概要・課題

- 深層学習による画像処理手法である「インスタンスセグメンテーション」はピクセル単位でボルダーを検出する手法である
- 物体の文脈を学習する新しい学習機構であるTransformerが登場しており、Transformerを沢山組み込んだ基盤モデルが登場している
- 基盤モデルを用いて惑星を解析する研究には前例がなく、既存のモデルでは検出性能が不十分
→ 複数の惑星のデータ作成
→ Fine-tuningすることでボルダー検出へ最適化

1-5 背景 - インスタンスセグメンテーション

- インスタンスセグメンテーションとは、画像内の個々の物体を識別し、マスクで分割するコンピュータビジョン技術
→ バウンディングボックスで物体を囲み、輪郭と主軸角度を取得可能
- MaskR-CNNの学習アーキテクチャ



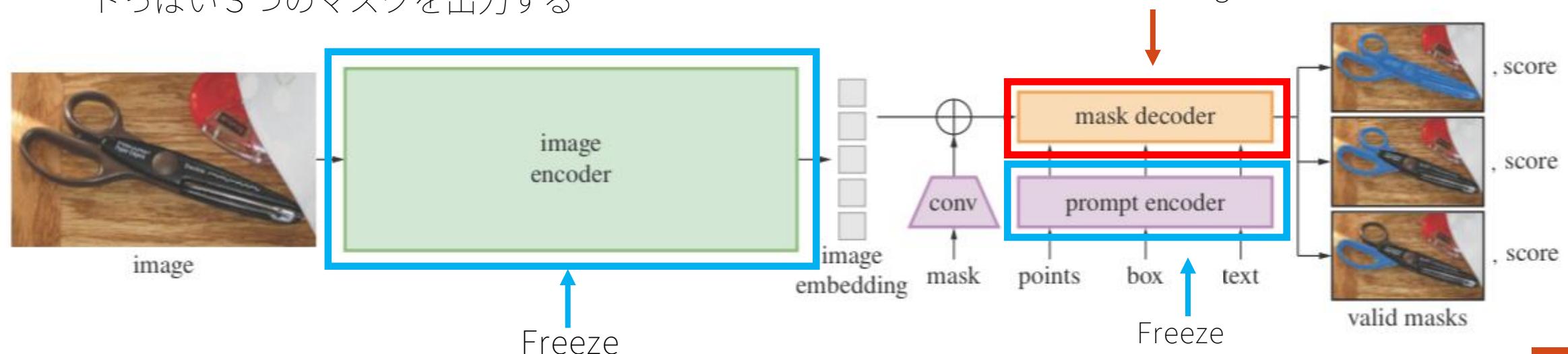
- Backbone: 特徴抽出ネットワーク。異なるサイズの物体を検出するための特徴マップを提供
- Region Proposal Network (RPN): 物体が存在する可能性のある領域を提案
- Region of Interest (RoI) Align: 分類、ピクセル単位のマスク、バウンディングボックスを予測

1-6背景 - 基盤モデルの概要

基盤モデルとは膨大なデータで事前学習された汎用的なAIモデルである。

今回の研究ではMetaが開発したSAM(Segment Anything)を利用している

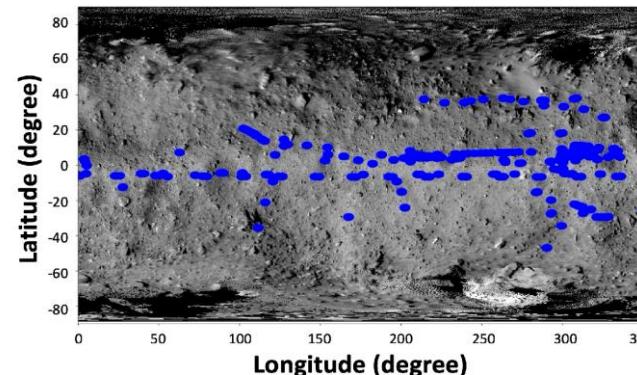
- 1100万枚と11億個のマスク情報で学習している
→Mask Decoderを少量学習することでボルダーに特化したモデルを作成
- モデルの出力にはどの位置に集中してほしいのかをユーザーが選択する必要があり、最もオブジェクトっぽい3つのマスクを出力する



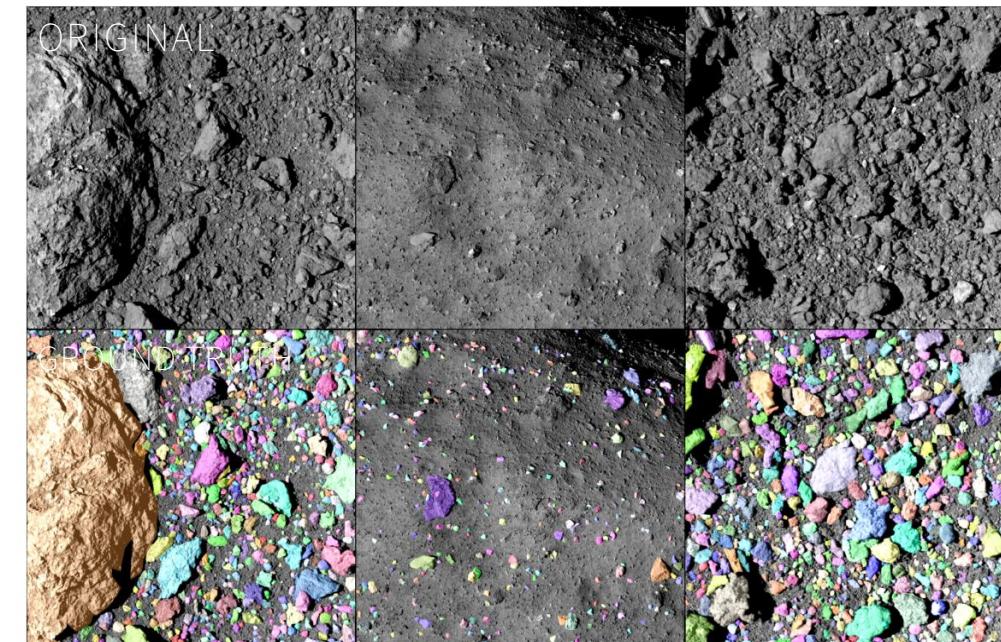
2-1 リュウグウデータセット

■ データセット概要

- はやぶさ2の光学カメラONC-Tが撮影した画像を利用 [3]
- 画像はISAS/JAXA Data Archives Transmission System (DARTS)から公開されている(下記URL参照)
 - https://darts.isas.jaxa.jp/pub/hayabusa2/onc_bundle/
- 228枚の画像から10ピクセル以上のボルダーを合計117,676個をアノテーション
- 1024 * 1024pixelの画像サイズ
- 高度5km以下で撮影した画像に限る
- 可視光領域であるv-band filterの画像を利用



データの撮影場所の分布



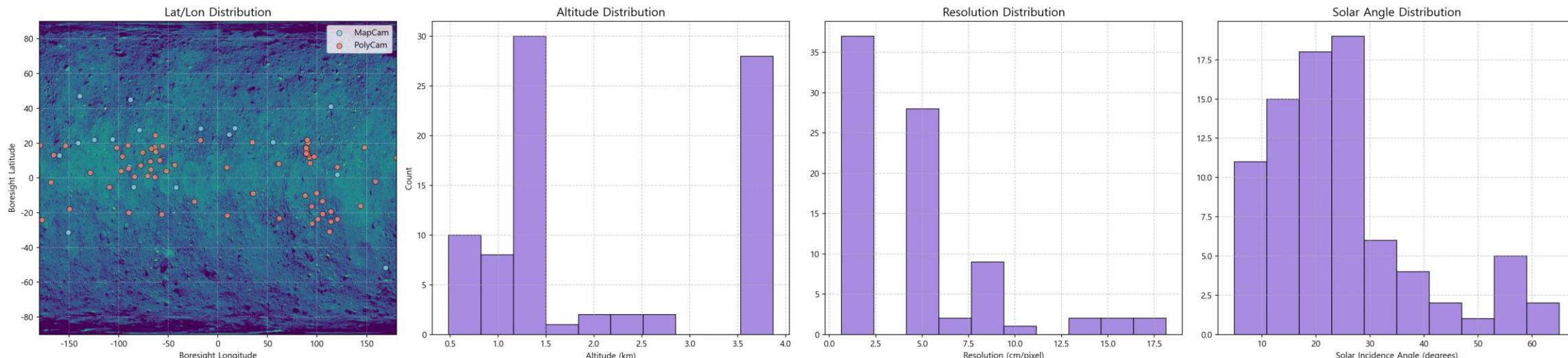
データセットのサンプル画像

2-2 ベヌーデータセット

■ データセット概要

- OSIRIS-REx Camera Suite (OCAMS)で撮影された画像を使用
- 画像はOSIRIS-REx Mission Data Archiveから取得
<https://sbnarchive.psi.edu/pds4/orex/orex.ocams/>
- 83枚の画像から、125,442個のボルダーアノテーションを付与
- 可視光領域であるv-band filterの画像を利用
- 高度、ボルダーの大きさに対し多様性を持つ

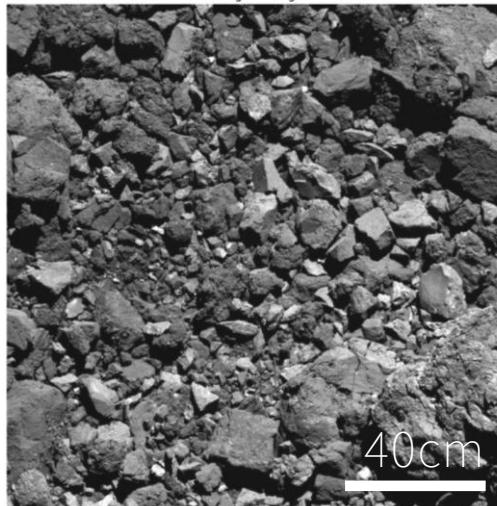
Bennu Dataset overview (Total n=83)



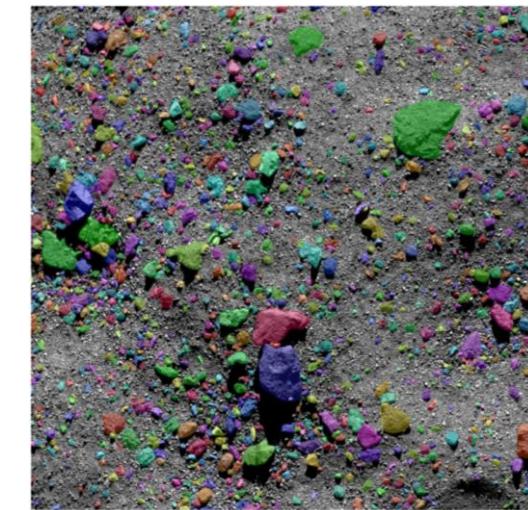
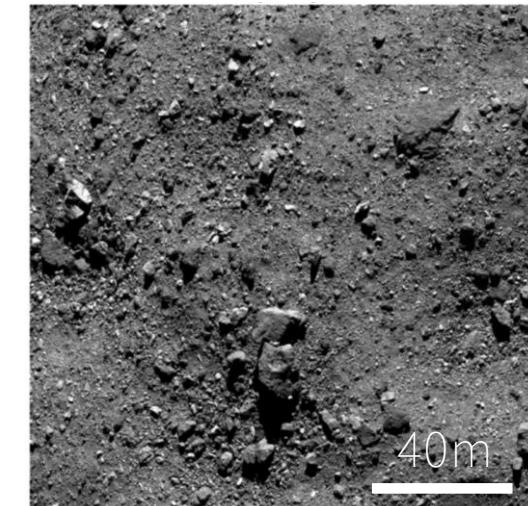
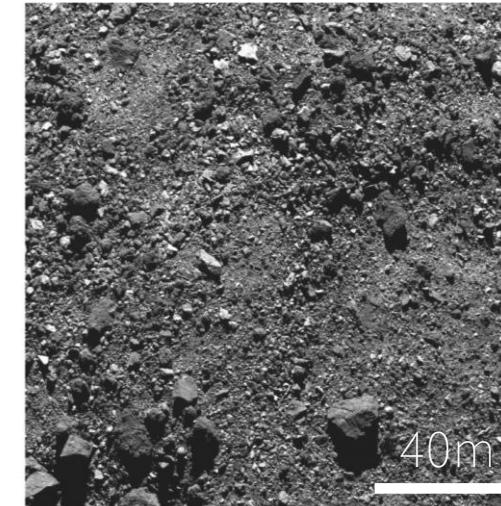
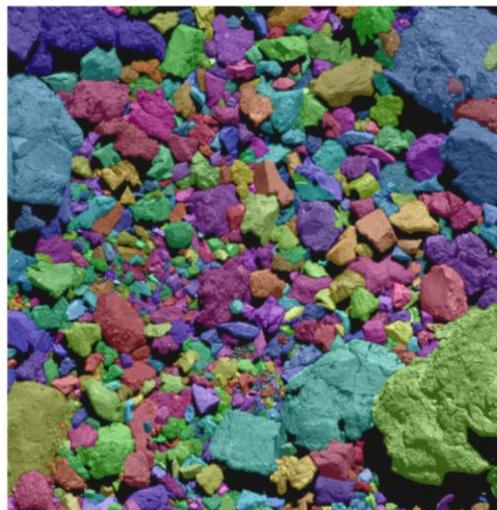
2-3 ベヌーデータセット

■ サンプルベヌーデータの可視化

Input



Ground Truth



2-4 データセットの前処理と拡張

先行研究により画像の大きさを減らすと精度が上がることが報告されている

■ データ拡張

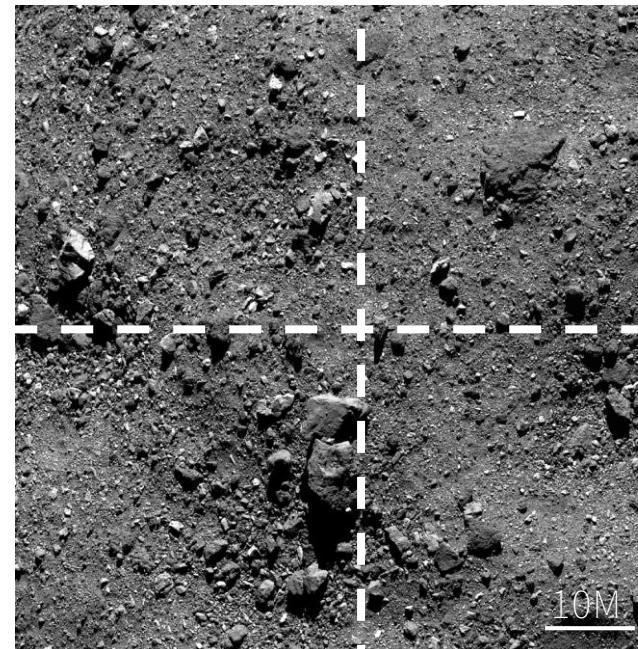
- 各画像を4分割
- ボルダーのアノテーション位置調整

■ 前処理

- Sigma contrast
→ ボルダー強調
- Wiener deconvolution
→ ぼやけ除去

■ 学習、テストセットの分割

- 全データセットからランダムに選択し、学習、テストセットを8:2の割合で分割した



データ拡張のサンプル画像

Ryugu Half データセット概要
画像サイズ: 512*512
全画像数: 897
全アノテーション数: 115,698
平均ボルダー数: 135

Bennu Half データセット概要
画像サイズ: 512*512
全画像数: 332
全アノテーション数: 129,594
平均ボルダー数: 390

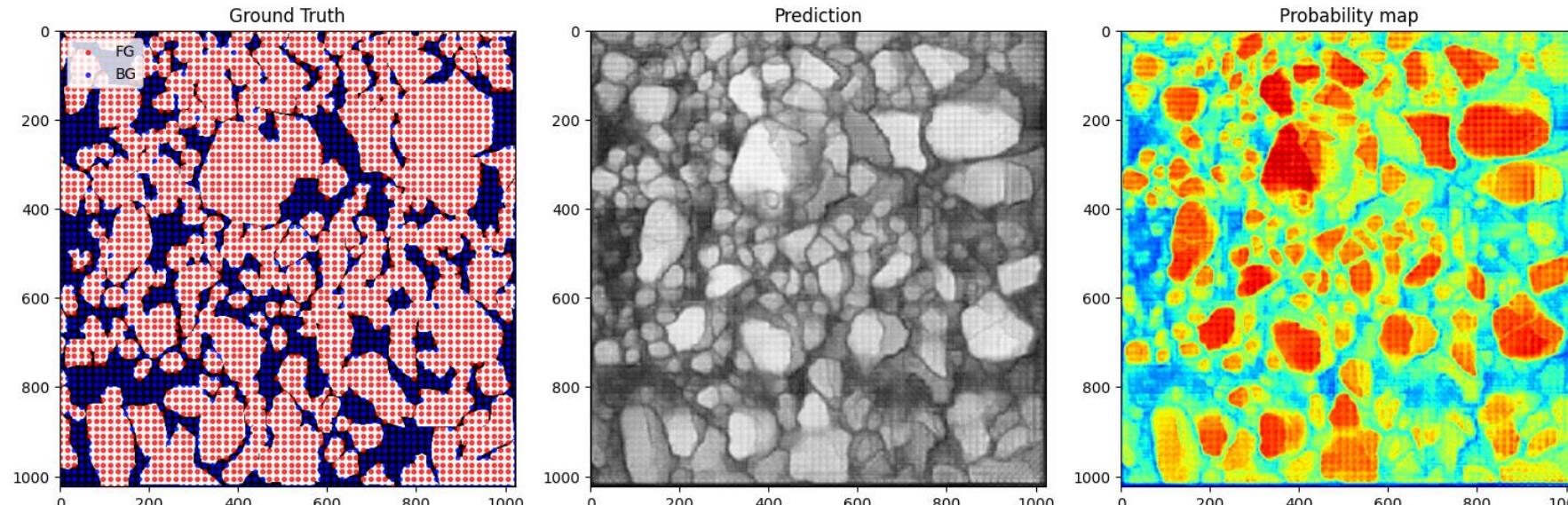
2-5 Fine-Tuning

SAMモデルを用いたボルダー検出の問題点

- モデル検出にはプロンプトが必要なため、正確に正解ラベルの位置を把握する必要がある
- 影情報もオブジェクトとして認識する

モデル入力・出力

- 評価と同じ環境を作るために各辺に64個の点をつけ、プロンプトとして利用
- ボルダーの居場所を確立マップとして出力する



2-6 口ス関数

■ Focal 口ス

- クラス不均衡を解決するために考案され、特にObject detectionやBinary segmentationでよく使われる
- 分類することが難しいターゲットに焦点を当てている

$$\text{Focal Loss}(p_t) = -\alpha(1 - p_t)^\gamma * \log(p_t)$$

■ Dice 口ス

- GTマスクと予測マスクの違いを測定
- 特に小さいかまだらなターゲットを扱うときに有効

$$\text{Dice Loss} = 1 - \frac{2|A \cap B| + \text{smooth}}{|A| + |B| + \text{smooth}}$$

■ モデルへのアップデート

$$\text{Total Loss} = \alpha * \text{Focal Loss} + \beta * \text{Dice Loss}$$

2-7 性能評価方法

■ 検出率

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

ここで, ‘正しく検出’ は $\frac{\text{Area of Intersection}}{\text{Area of Union}}$ ($\text{IoU} > 0.5$) を意味する

■ Overall IoU

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

■ Overall Accuracy

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

F1スコアを起用しなかった理由

- 画像単位にボルダーの形を獲得するより、個数を正確に得ることが重要
- ボルダーが小さいため、誤検知と評価されても正しく検出している場合がある

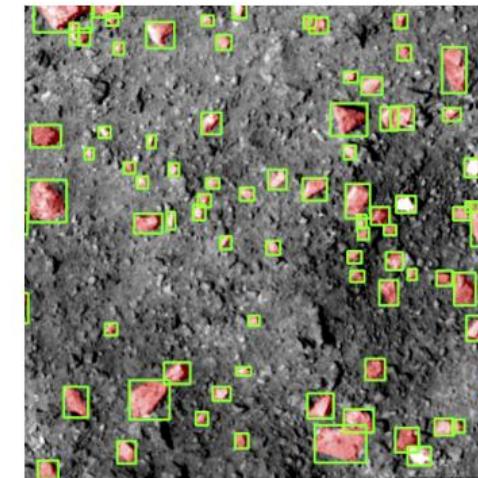
3-1 結果 – テーブル

■ Recall, IoU, Overall accuracy (OA)によるResnet-50とSOD-Formerの定量比較

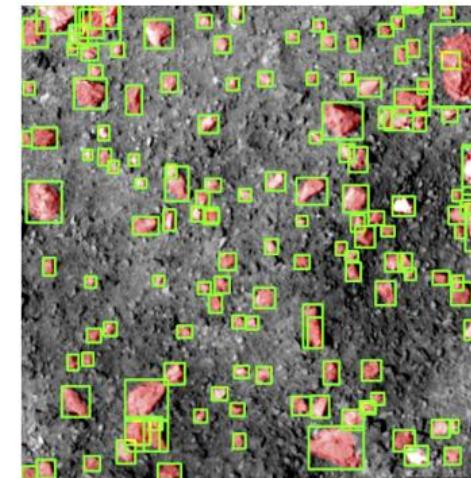
Model	Recall (%)	IoU (%)	OA (%)	# boulder
Ryugu Dataset				
SOD-Former	59.55	68.48	84.32	127
SAM	57.05	55.36	85.30	202
FT-SAM	68.25	62.01	87.25	372
Bennu Dataset(Trained with only Ryugu dataset)				
SOD-Former	22.04	44.89	61.45	118
SAM	56.59	66.68	78.14	413
FT-SAM	75.43	81.60	87.55	781

3-2 結果 - 画像(Ryugu)

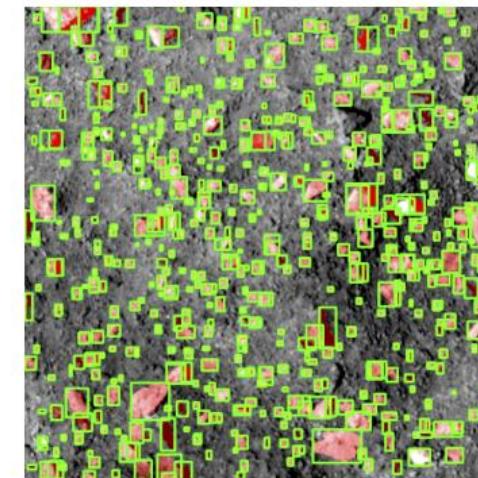
■ リュウグウで学習を行い、リュウグウで評価した結果



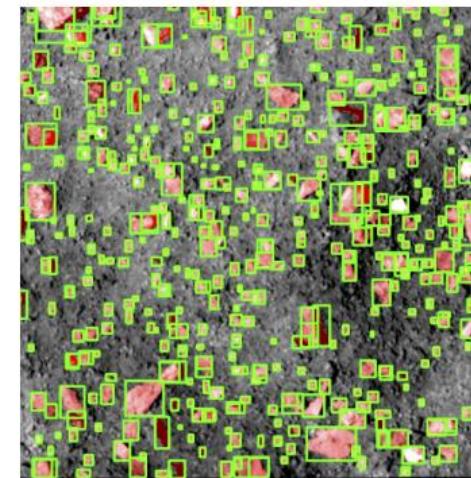
Original Image



SOD-Former



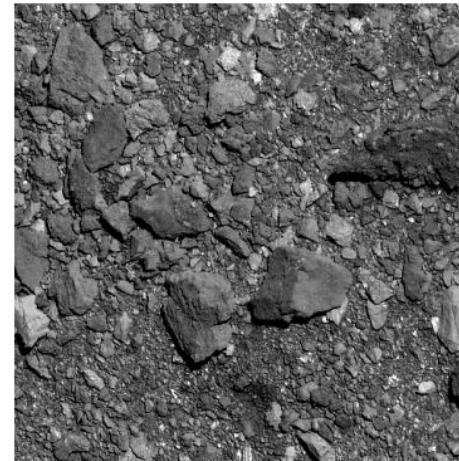
Base-SAM



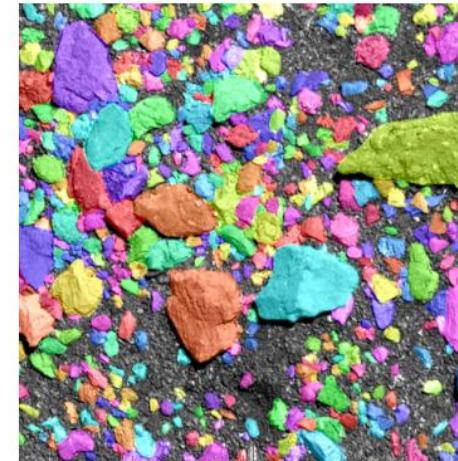
Finetuned-SAM

3-3 結果 – 画像(Bennu)

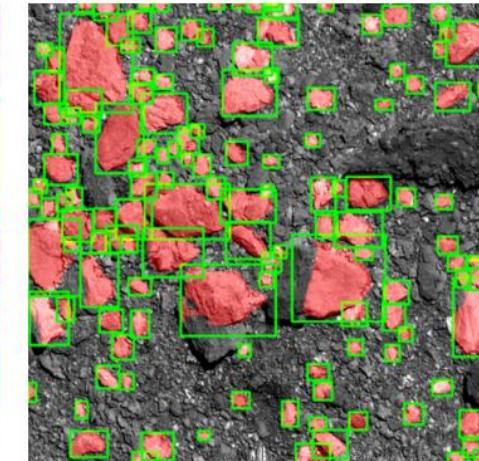
■ リュウグウで学習を行い、ベヌーで評価した結果



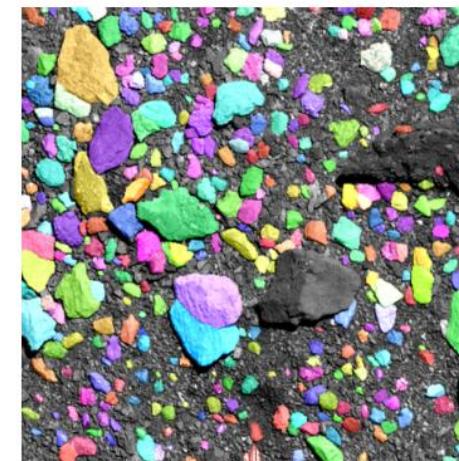
Original



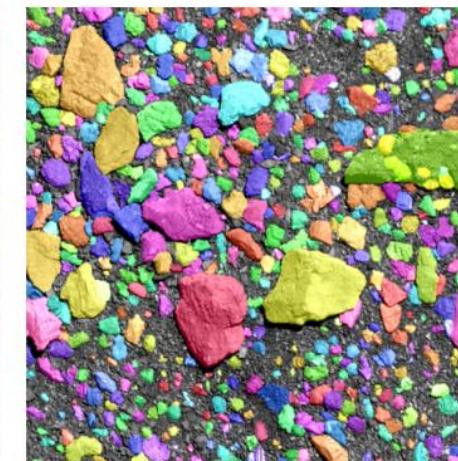
Ground Truth



SOD-Former

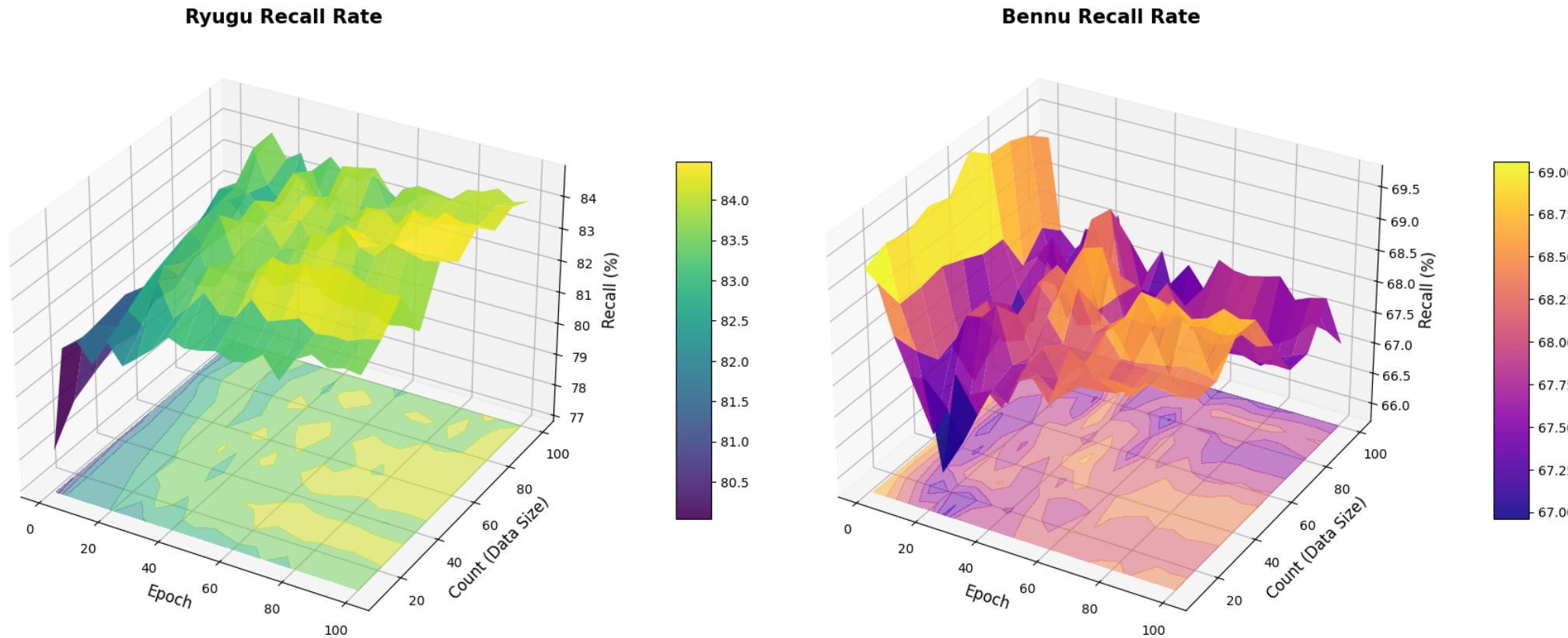


Base-SAM



FineTuned-SAM

3-4 結果 – Finetuningの最適学習枚数、エポックの調査



考察 & 今後の展望

■ 考察

本研究では他の小惑星でも使える基盤モデルを実装することを目的として**SAM**をFine-tuningし、小惑星リュウグウ上のボルダーの検出性能を行った。結果として、従来の深層学習モデルは新しい小惑星に対して汎化性能が低下した反面、基盤モデル**SAM**をファインチューニングしたモデルは、未知の小惑星ベヌーで圧倒的な汎化性能を示した。これは、特定の学習データに依存せず「物らしさ」を捉える基盤モデルが、深宇宙探査においても有効であることを示唆する。

■ 今後の展望

- 基盤モデルのFine-tuningに最適したロス設計、画像枚数、エポックなどの調査
→機械学習に詳しくない科学者が素早く適用できる環境実装
- 小惑星リュウグウ、ベヌーの数密度分布、累積サイズ分布の全休規模での調査
→二つの小惑星間の違いの解明