

領域分類可能な3次元物体 生成モデル

B4 中須賀 大輝

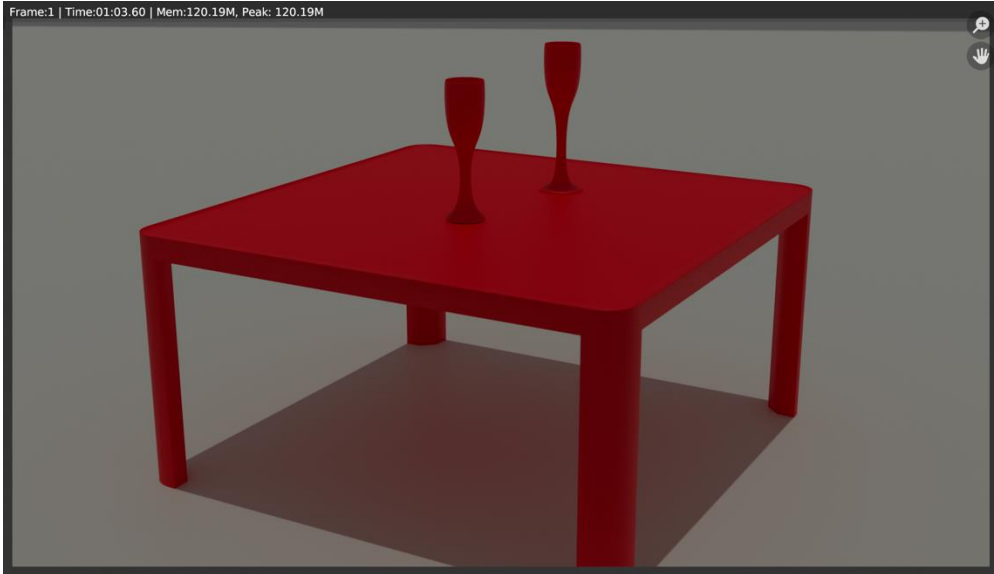
研究背景

- 近年，多視点画像から3D Gaussian Splatting[Kerbl+]を用いて3次元物体を作る研究が注目されている
- 1つのシーンを作るのに必要な多視点画像数は100~200枚
本研究でシーンとは空間に映る情報(座標，色，透明度，個々の物体の形など)全体を纏めたデータを表す

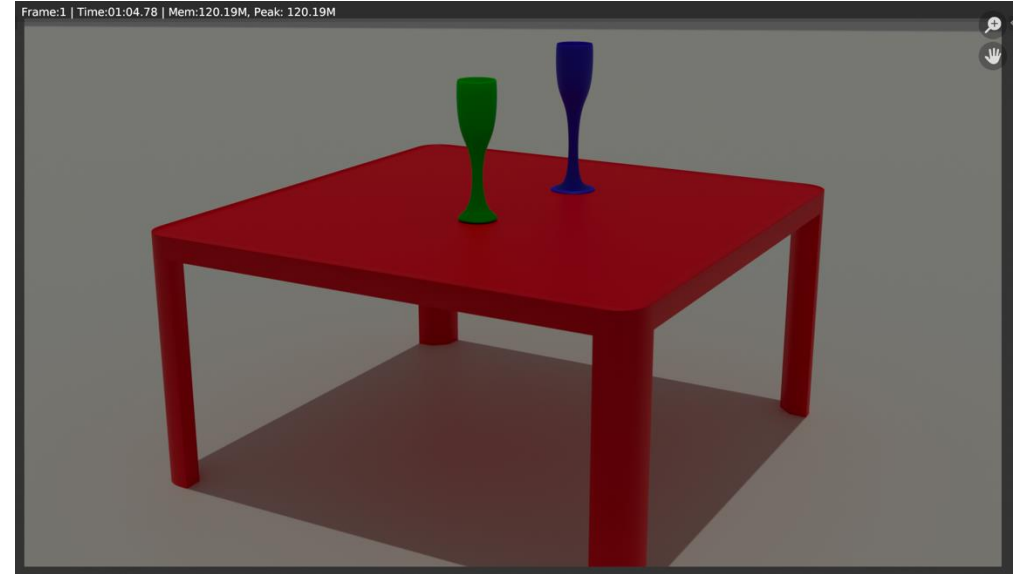


研究目的

- ・ 先行研究: テキストを用いて部分的に3Dシーン取得可能
- ・ 本研究: テキストなどを介さずに自動で3Dシーンを取得



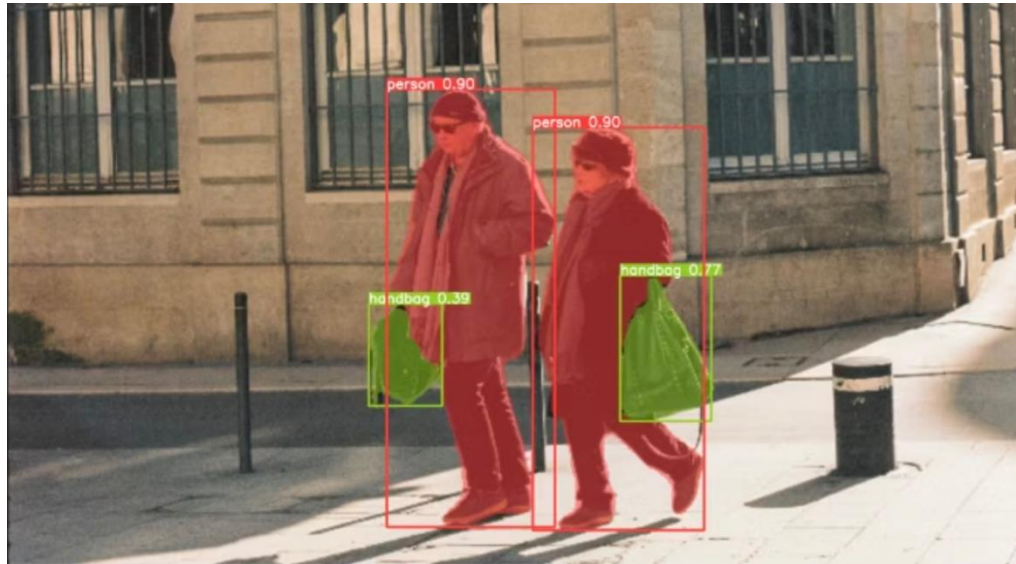
現状



目標

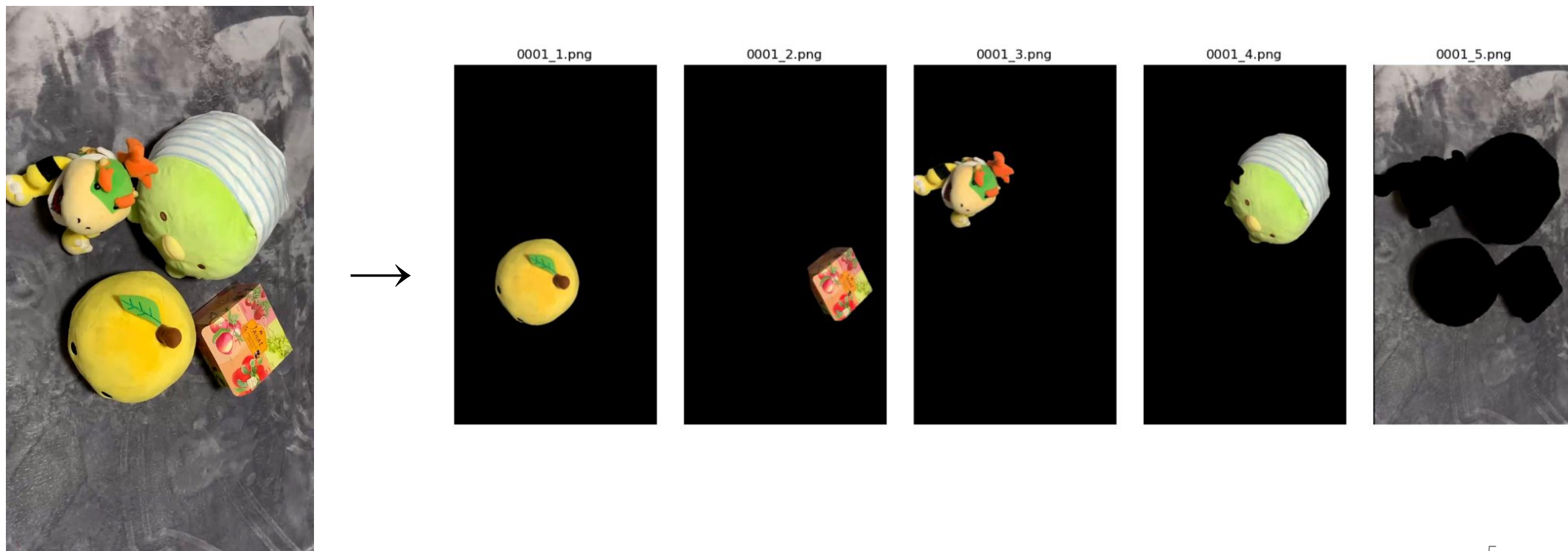
領域分類

- ・ 画像や動画から情報(下図では人とバック)を取得
- ・ 3次元物体生成モデルと併用



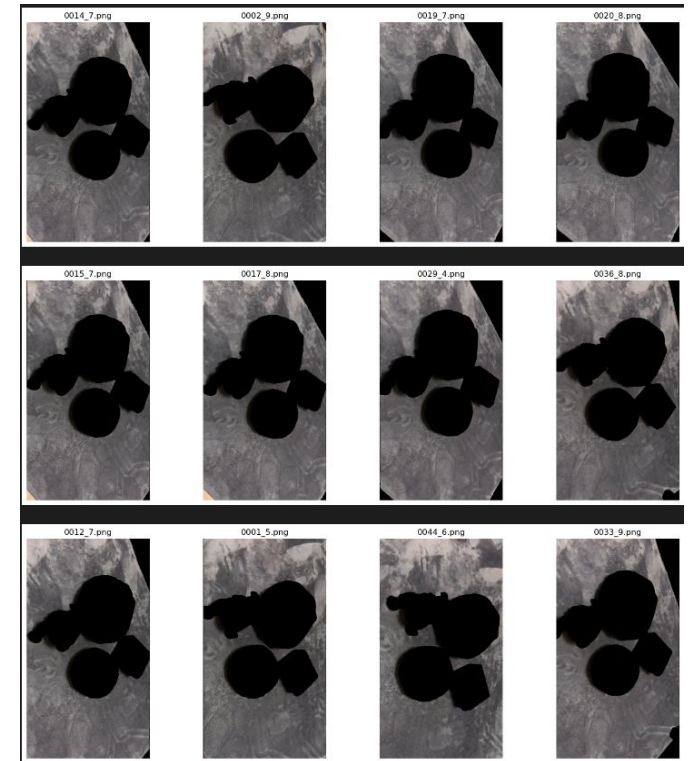
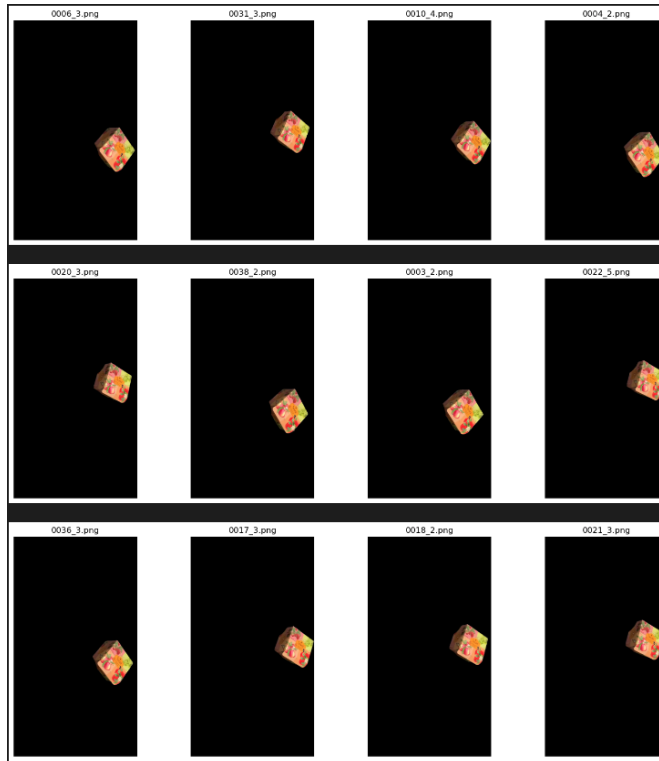
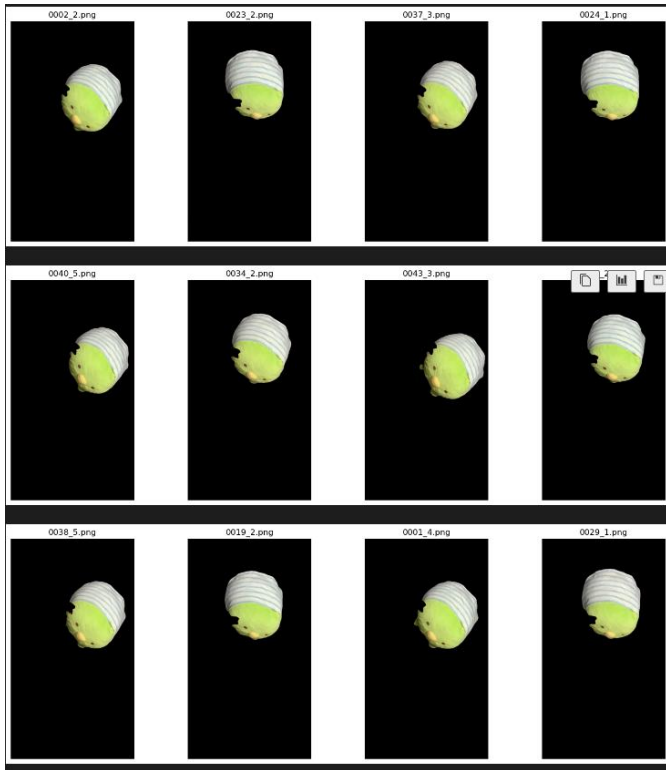
提案手法

Segment Anything Model(SAM) [Kirillov+]を扱い全
多視点画像に対して物体を分割



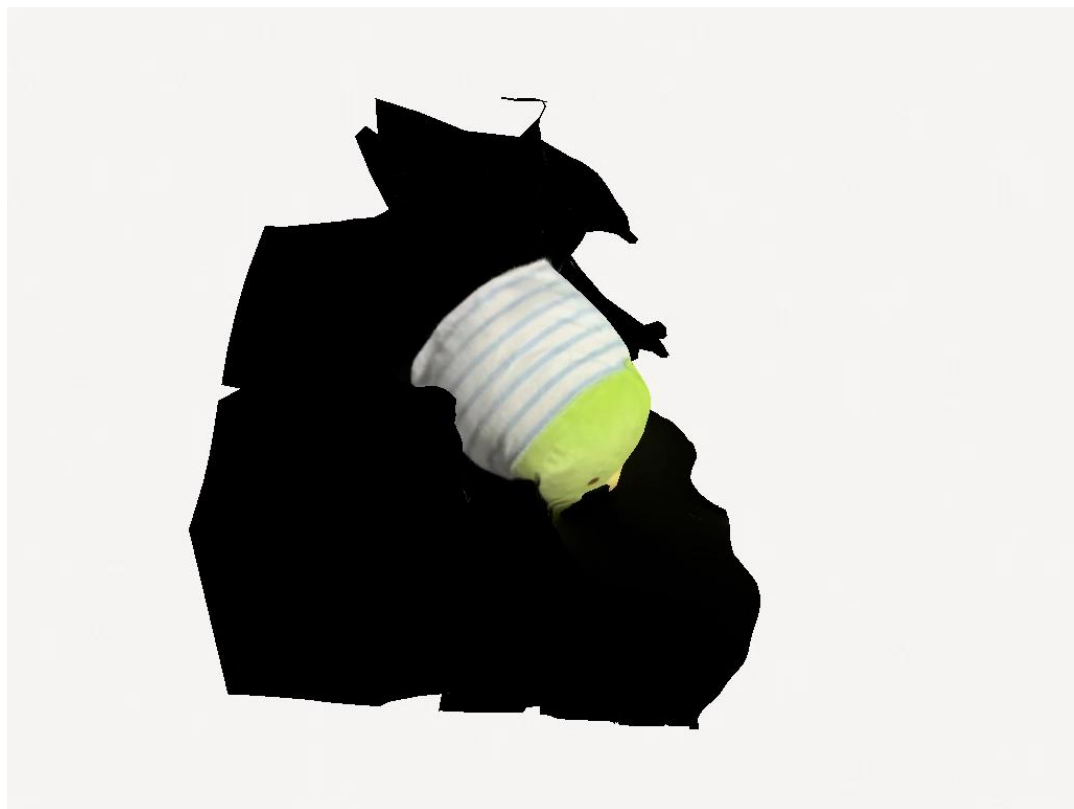
提案手法

- ・ クラスタリングを行う
- ・ 3つのクラスタリングの手法を試行し比較



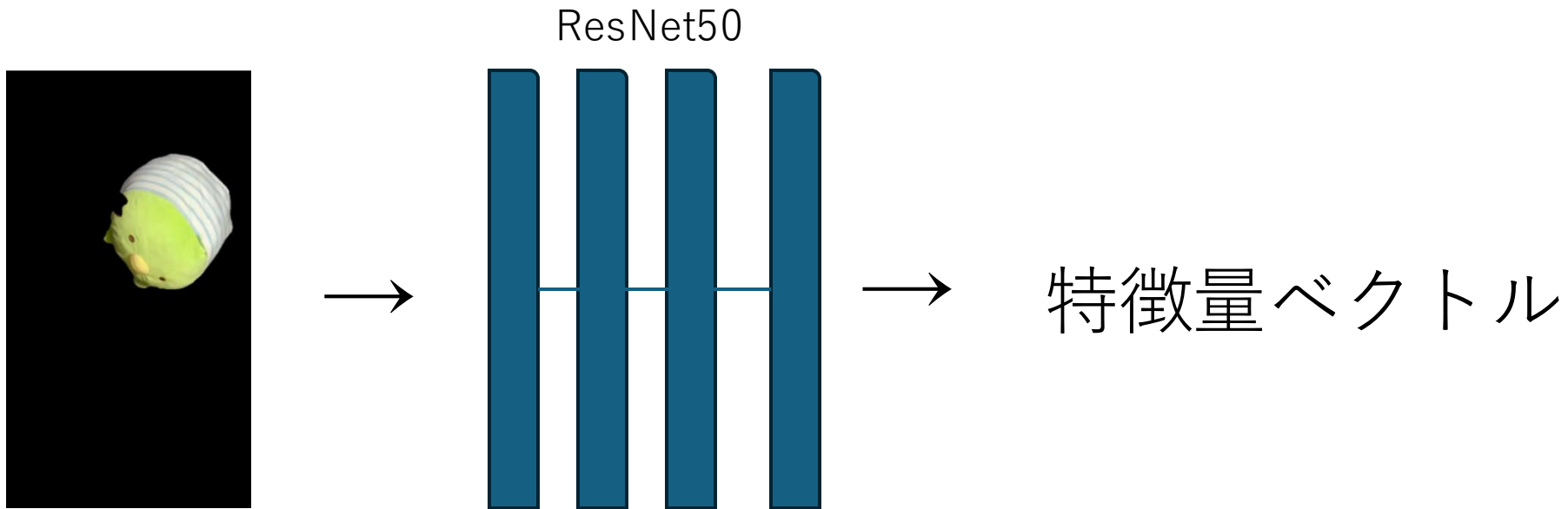
提案手法

3次元物体生成モデルを使用し3D物体を生成



クラスタリングの前処理

- クラスタリングのため画像をベクトル化
- ベクトル化にはResNet50 [Kaiming+]というCNNを使用



クラスタリング手法

- 類似度閾値ベースクラスタリング
画像間の \cos 類似度を使用
- K-means ベースクラスタリング
K-means [Raschka+]を使用
- HDBSCAN ベースクラスタリング
HDBSCAN [Raschka+]を使用

精度指標:正答率

クラスタ間で最適マッチング[穴井+]
行い,

$$\text{正答率} = \frac{T}{N}$$

(正しく分類された画像数 T , 画像の総数 N)

これでクラスタリングの精度を評価

テストとしてシーン1,シーン2の2種類の多視点画像を用意



シーン1.



シーン2.

結果

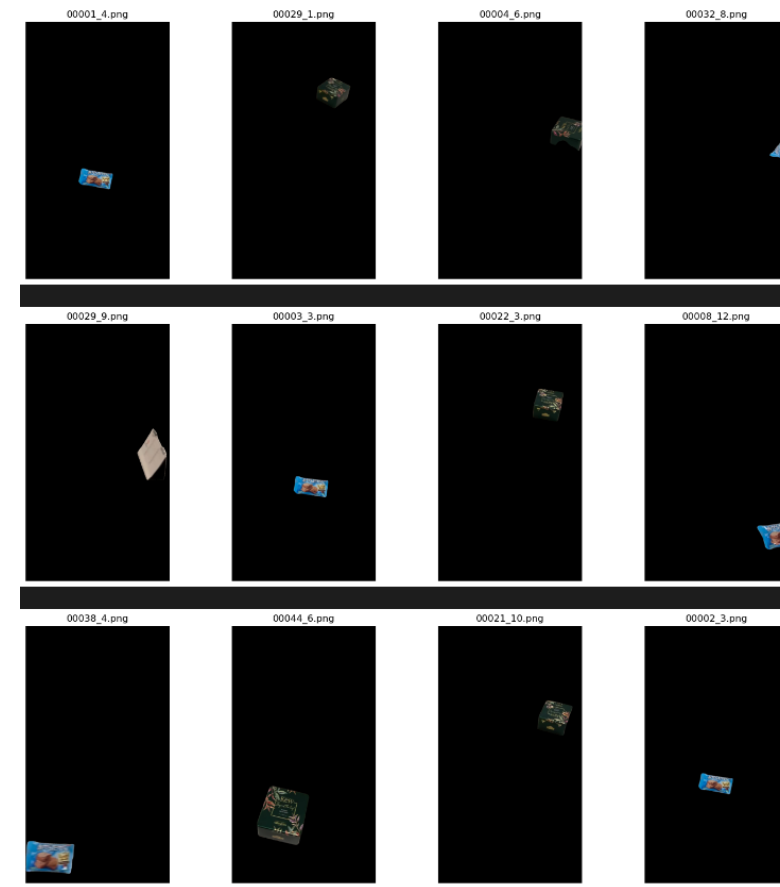
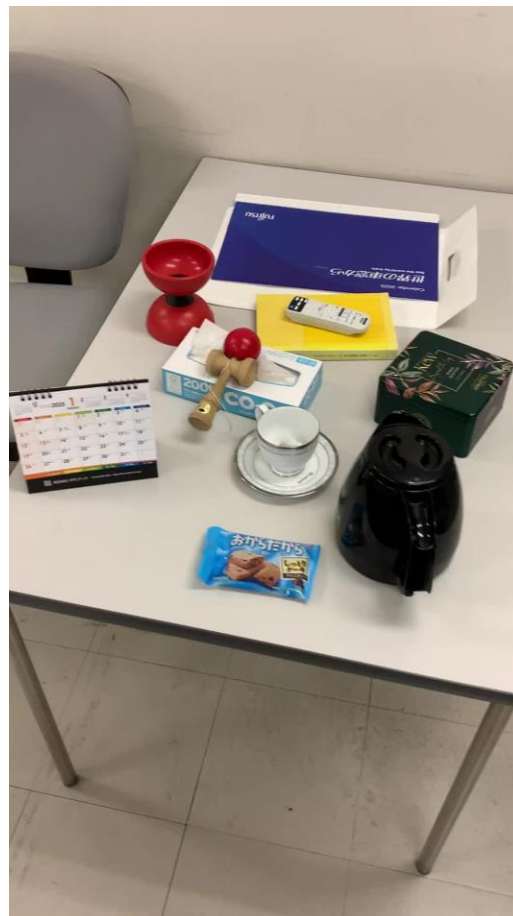
- ・最も良い正答率なのはK-meansベース
- ・より複雑なシーン1の正答率が低い

	シーン1	シーン2	平均
類似度閾値ベース	57.4%	95.3%	76.35%
K-meansベース	65.2%	96.7%	80.95%
HDBSCAN	42.6%	87.5%	65.05%

シーン1の正答率が低い原因

シーン1では
パラメータの自動調整の
精度に難あり

→手動でパラメータ調整が
必要



課題

- 左図の画像に対しては補間が必要
- 右図では背景が黒なため背景と物体が一体化
→ 前処理として背景部分の削除が必要
- パラメータの細かい調整



参考文献

[Kerbl+] 3D Gaussian Splatting for Real-Time Radiance Field Rendering(2023)

[Kirillov+] Segment Anything(2023)

[Kaiming+] Deep Residual Learning for Image Recognition(2015)

[穴井+]今日から使える組み合わせ最適化

[Raschka+] **Python**機械学習プログラミング：達人データサイエンティストによる理論と実践

Appendix

- 類似度閾値ベースクラスタリング

cosine距離で一度集めたものを再度メドイドを使って閾値で合成してる． 閾値はhold-out検証で決めている

- K-means ベースクラスタリング

Kはシルエットスコア(凝集度と乖離度)が高いものを選んでいる．

- HDBSCAN ベースクラスタリング

予めエルボー法を用いてPCAを行っている．