

第51回海洋開発シンポジウム

視覚的言語記述に基づく沿岸インフラの抽出

Segment Anything Model 3 (SAM3) を用いたゼロショット推論の検証

増田 和輝¹・田中 陽二¹・河田 晃靖¹・下園 武範²

¹ 東洋建設株式会社 総合技術研究所

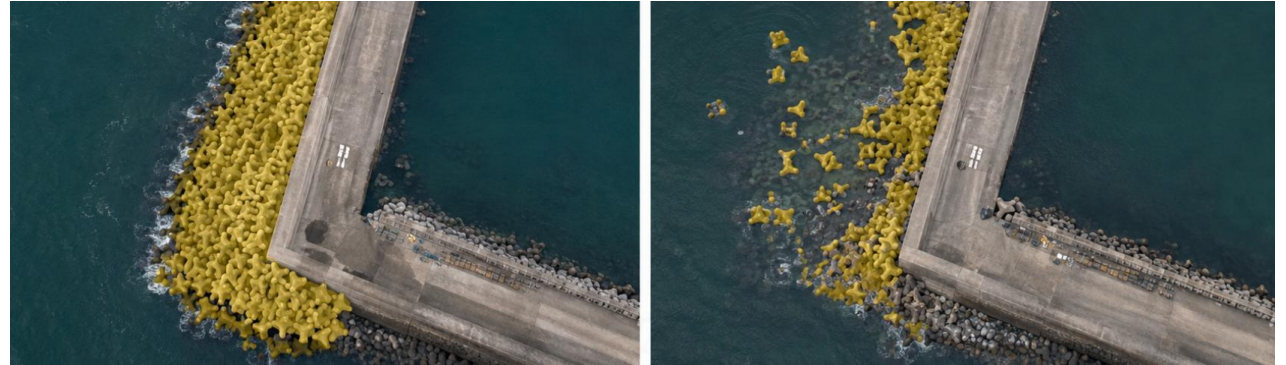
² 東京大学 工学系研究科社会基盤学専攻 教授



沿岸インフラの広域モニタリング

- 防波堤や消波ブロック等の港湾インフラの継続的な維持管理は防災上重要
- 高解像度衛星・航空画像の利用が進展
広域・高頻度のモニタリングにおける目視判別は労力が大きい
- AIによる自動モニタリングが期待される
⇒ “セグメンテーション” による領域比較

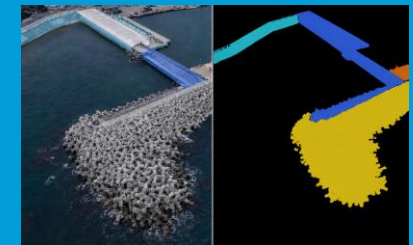
例：災害時の迅速な被害量の確認など



学習データの用意

膨大なアノテーションコスト

- 対象構造物ごとに「ピクセル単位の精緻な教師データ」が大量に必要



課題

対象ごとに大量の精緻な教師データが必要

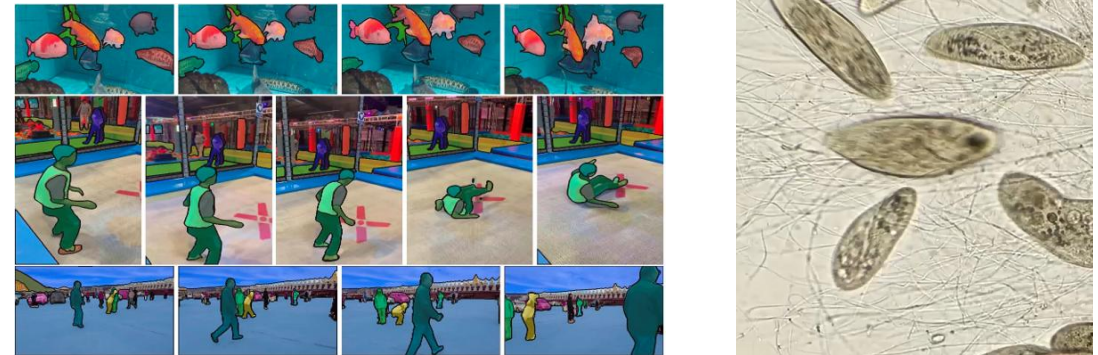
着眼点

新たに学習せずにオープンで公開されているAIモデルをうまく活用しよう

Segment Anything Model

Meta社が開発したセグメンテーションモデル

大量のセグメンテーションデータを学習
⇒汎用的な物体の区分分けが可能
⇒セグメンテーションの基盤モデル



SAMのデモ

<https://ai.meta.com/blog/segment-anything-2/>

学習済みモデルをダウンロード可能

2023

SAM

点・矩形・マスクなどの
プロンプトから領域抽出

2024

SAM2

画像から動画へ
追跡と時系列補正

2025

SAM3

言語概念・例示画像から
該当インスタンスを抽出

基盤モデル：大量のデータを用いて事前に学習され汎用性の高い大規模なAIモデル

検討内容

セグメンテーションの基盤モデルであるSAM3を用いて
自然言語の指示（プロンプト）から沿岸インフラをゼロショット抽出

基盤モデル：大量のデータを用いて事前に学習され汎用性の高い大規模なAIモデル

ゼロショット：追加の学習や具体的な正解例を一切与えず、指示のみで推論する手法

本研究の問い

基盤モデルのAIは一般的なことは大量に学習が行われている
⇒土木の専門知識をどの程度が認識できているのか？

本研究の目的

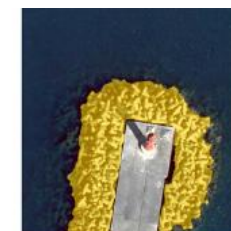
- ・ 基盤モデルが沿岸インフラをどの程度判別できるかの定量的評価
- ・ 精度よく判別するためのプロンプトの改良方法を提示



プロンプト
“Block”
“Tetrapod”



基盤モデル
SAM3



検証に国土地理院のGSIデータセット（公開データ）を利用

「CNNによる地物抽出用教師データセット」、地上画素寸法20cm級の空中写真画像を対象として、道路や建物など様々なカテゴリにアノテーションされているセグメンテーションの学習用データセット

国土地理院 GSIデータセットより

https://gisstar.gsi.go.jp/gsi-dataset/index_ja.html



オリジナル画像

ラベリング画像

Water

水域

海面・港湾内・河口域

Breakwater

防波堤

ケーソン堤・護岸等の線状構造物

Block

消波ブロック テトラポッド等の異形ブロック群

プロンプト

専門語
視覚記述
空間文脈

SAM3

テキスト指示のみ
Fine-tuningなし

評価

IoU / F1
Precision / Recall

- IoU：正解領域と予測領域の一致度.
- Precision：誤検出の少なさ.
- Recall：検出漏れの少なさ.
- F1：PrecisionとRecallのバランス.

専門語をそのまま入れるのではなく，形状・材質・状態・場所へ分解して言語化
以下の3つの論理的意図をもとにプロンプト群を設計した

※詳細な実験ケースは論文をご確認ください

A _____

専門用語

Breakwater

防波堤

Tetrapod

テトラポッド

Wave-dissipating block

消波ブロック

モデル内に専門概念が
存在するかを確認.

B _____

視覚的特徴

long straight concrete

長い直線上のコンクリート

randomly piled concrete

ランダムに積み上げられたコンクリート

interlocking concrete shapes

連結するコンクリートブロック

形状・材質・テクスチャにより
認識を誘導.

C _____

空間的文脈

extending into the sea

海に向かって伸びる

along the coast

海岸沿いに

on the shore

海岸に / 岸辺に

類似構造物との混同を周辺環境
で抑制.

Best Prompt

“Water”

最も広義な語が最良。Sea / Oceanは
港湾内・河川を拾いにくい。

F1-score

0.788

Precision 0.981 / Recall 0.659

検討プロンプト

日本語訳

"Water"

水

"Sea", "Ocean"

海, 大洋

"Water surface area"

水面領域

国土地理院の研究ではF値0.8以上を高精度としてAIモデルの学習を実施

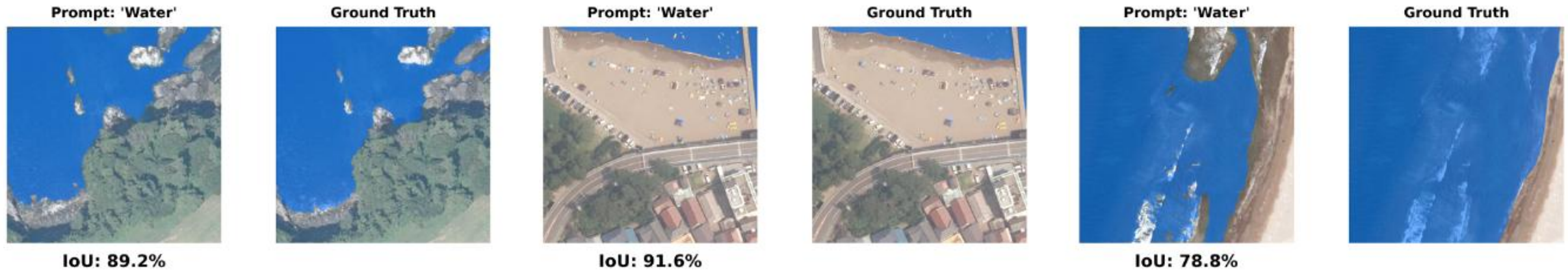


図-1 Water（水域）におけるプロンプト記述の抽出結果の一例

水域は一般的な用語として認識される一方、砕波や水際の境界では検出漏れが発生する。

Standard Prompt

“Breakwater”

F1-score

A: 専門用語

0.332

Precision 0.361 / Recall 0.307

Best Prompt

海に突き出た直線状のコンクリート建造物

“Linear concrete structure extending into the sea”

F1-score

B: 視覚的特徴

C: 空間的文脈

0.748

Precision 0.618 / Recall 0.946

Prompt: 'Breakwater'



IoU: 56.6%

Prompt: 'Linear concrete structure extending into the sea'



IoU: 96.0%

Ground Truth



Prompt: 'Breakwater'



IoU: 0.0%

Prompt: 'Linear concrete structure extending into the sea'



IoU: 62.8%

Ground Truth



図-2 Breakwater (防波堤) におけるプロンプト記述の差異と抽出結果の一例

消波ブロックも含めて防波堤の一部とも考えられるが、本研究はGSIデータセットのラベルを正とした

- 形状と配置（どこにあるか）を明示すると、線状構造物としての抽出が安定。
- 一方で、岸壁、道路、建屋などの類似構造物も拾いやすく、誤検出が多くなる傾向。

Standard Prompt

“Blocks”

F1-score

0.017

Precision 0.312 / Recall 0.009

Best Prompt

海の中にランダムに積み上げられたコンクリートのがれき

“Randomly piled concrete debris in the sea”

F1-score

0.799

Precision 0.741 / Recall 0.867

B:視覚的特徴

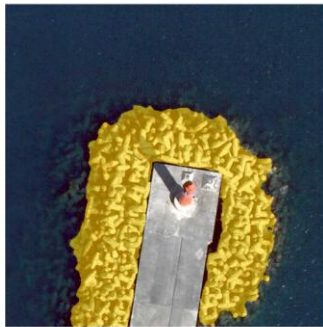
C:空間的文脈

Prompt: 'Blocks'



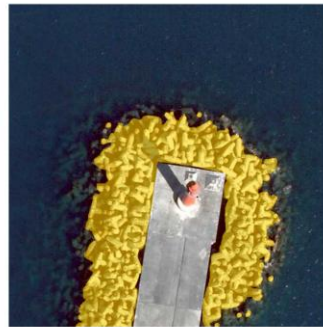
IoU: 0.0%

Prompt: 'Randomly piled concrete debris in the sea'



IoU: 85.5%

Ground Truth



Prompt: 'Blocks'



IoU: 16.7%

Prompt: 'Randomly piled concrete debris in the sea'



IoU: 8.2%

Ground Truth

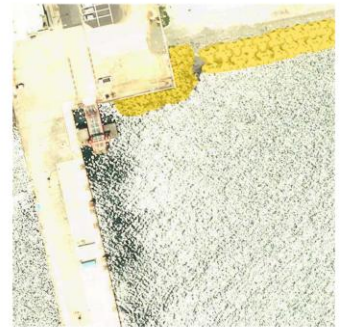
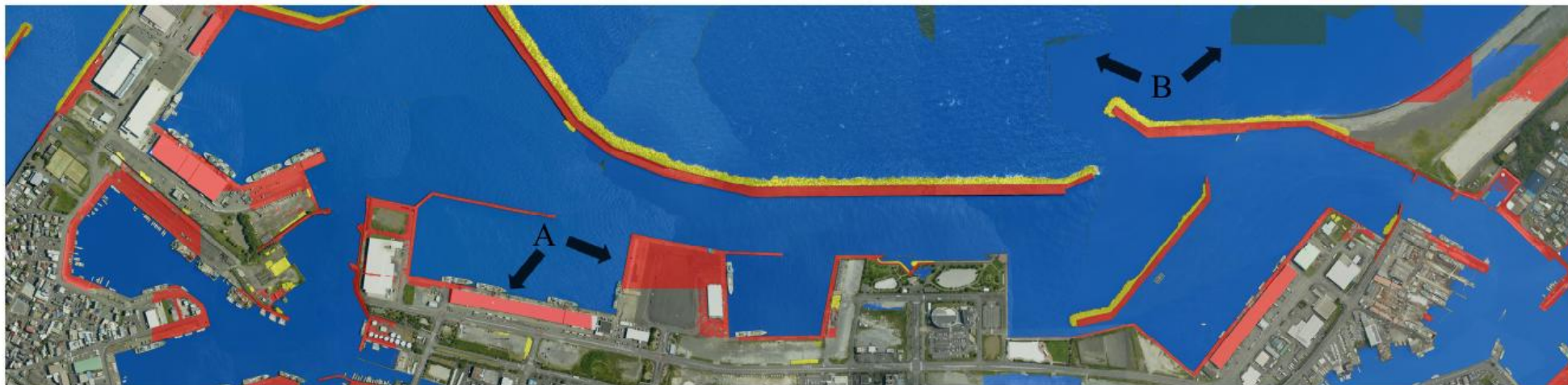


図-3 Block（消波ブロック）におけるプロンプト記述の差異と抽出結果の一例

- 個々ではなく、テクスチャとして「海中の不規則なコンクリート集合体」として記述すると認識しやすい。
- 乱反射が集合体として誤検出されるケースも存在した。

実フィールドを想定した検証：焼津漁港の広域抽出

静岡県「VIRTUAL SHIZUOKA」より、焼津漁港周辺のオルソ画像（20cm/pixel）を使用



- Water: 'Water'
- Breakwater: 'Linear concrete structure extending into the sea'
- Block: 'Randomly piled concrete debris in the sea'

今回の検討で
最良のプロンプトで実施

図-4 焼津漁港における抽出結果

処理条件

1024×1024 pxパッチに分割し、50%重複で推論
重複画素は最大値統合方式で採用

確認された課題

建屋・護岸部など類似構造物の過剰検出
オルソ画像の継ぎ目・ノイズによる誤認

結論：専門用語を「見た目の言葉」に翻訳する

- 専門用語のみでの抽出は不安定.
- 形状・材質・テクスチャ・空間文脈を含む記述で精度が向上
- AIの学習を新たにしなくても、公開されている基盤モデルを工夫次第で特定の物体を抽出できる



学習を必要しないため、ユーザーに比較的素早く提供でき、他の海岸・港湾施設にも展開しやすい。

