

SAM2と**CLIP**によるセマンティックセグメンテーション

元画像



オーバーレイ（同じCLIPクラスは同色）



セマンティックセグメンテーション：画像の区分け

検出されたマスク統計（上位10マスク、CLIP分類付き）：
総マスク数：106

マスクID	面積	予測IoU	安定性	CLIPクラス	信頼度
0	33644	0.980	0.953	grass	0.543
1	33141	0.970	0.946	animal	0.344
2	27391	0.943	0.956	wall	0.375
3	11498	0.934	0.976	animal	0.321
4	4337	0.832	0.962	clothes	0.866
5	3851	0.969	0.958	mountain	0.342
6	3329	0.861	0.947	wall	0.219
7	2074	0.957	0.963	ceiling	0.126
8	2059	0.961	0.958	wall	0.114
9	2007	0.974	0.982	wall	0.218

CLIP：画像分類。ここでは「区分け」ごとの分類を行っている

概要と仕組み

- 画像内の全オブジェクトを自動検出し、30種類のカテゴリに分類

1. SAM2 (Segment Anything Model 2)

- Meta AIが開発したセグメンテーションモデル
- 階層的Transformerによる高精度な境界検出を特徴

2. CLIP (Contrastive Language-Image Pre-training)

- OpenAI開発の画像-テキスト対応学習モデル
- 4億組の画像-テキストペアで事前学習済み
- 検出された区分け領域について、プログラムでは、30カテゴリに自動分類（30の中身は変更可能）

処理の流れ

- **ステップ1: SAM2による自動マスク生成** → グリッド状に配置された点から各オブジェクトの境界を検出
- **ステップ2: CLIPによるゼロショット分類** → 各マスク領域を30カテゴリ（壁、建物、空、人、車など）に分類
- **ステップ3: カテゴリ別色分け可視化** → 同じカテゴリのオブジェクトを同じ色で表示

活用場面

- シーン理解
- 構成要素の把握
- データセット作成の前処理（アノテーション作業の効率化）
- 自動運転・ロボティクス（環境認識）
- 異常部分の識別