視覚言語モデルと商品デザイン知識の組み合わせを用いたプロンプト生成による物体遮蔽に頑健なセグメンテーションの検討

鳥取大学 Tottori University
中村琉聖　米田駿介　井上路子　西山正志 (鳥取大)

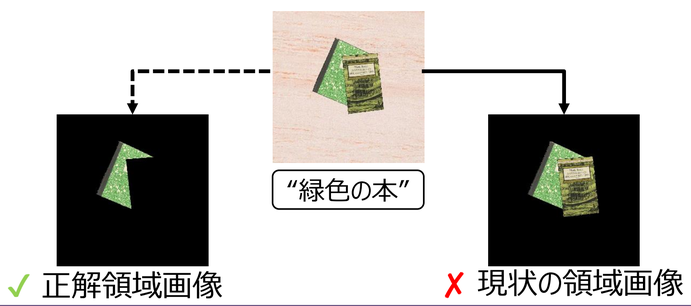
目的

プロンプト生成における試行錯誤の手間を省きつつ物体遮蔽に頑健なセグメンテーション手法を設計

● EVF-SAM[Yuxuan+,2024]：テキストプロンプトを用いることで物体遮蔽が生じていない場合に高精度なセグメンテーションが可能

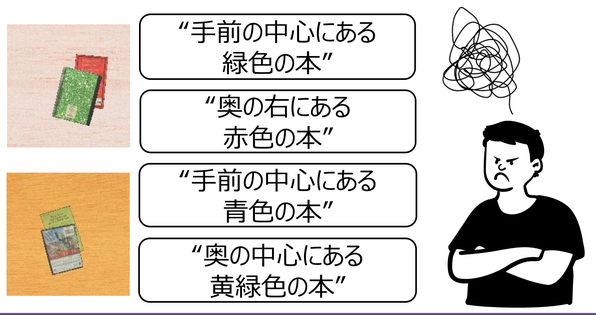
✗ 課題1

物体遮蔽が生じた場合に
位置情報を含まない
テキストプロンプトを与えると
セグメンテーション精度が低下



✗ 課題2

毎回の推論時に位置情報を含む
テキストプロンプトを人手で作るには
試行錯誤の手間が発生



提案手法

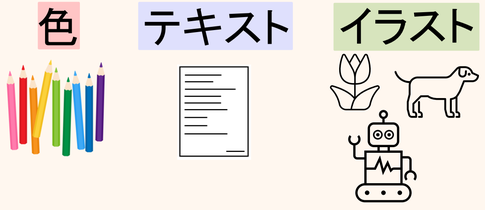
事前準備として遮蔽の影響を受けにくいテキストプロンプトデータベースを自動作成

アイデア①
商品デザイン知識に基づきVQAに与える質問文を設定

● 常に工夫が施され人間が直感的に認識しやすい
● 遮蔽が生じた場合にも影響を受けにくいと考えられる

商品デザイン知識

色　テキスト　イラスト



S1. 商品デザイン知識に基づき質問文を設定

“Please answer the color in the object.”
“Please answer whether or not there are texts.”
“Please answer whether or not there are illustrations.”

商品画像

アイデア②
視覚言語モデルを用いて商品デザイン知識に注目したテキストプロンプトを自動生成

● VQA: InternLM[Zhang+,2024]
● LLM: ChatGPT-4o
説明の順番と表現を統一

S2. VQAを用いてテキストプロンプトを自動生成

“It is a ①red book cover ②with white and black text. There are no ③illustrations on the cover.”

“It is a DVD ①cover with the color of white. There ③are illustrations ②and letters on it.”

S3. LLMを用いてテキストプロンプトを調整

“It is a ①red book cover. There is white and ②black text. ③There are no illustrations.”

“It is a ①white DVD cover. There is ②text. ③There are illustrations.”

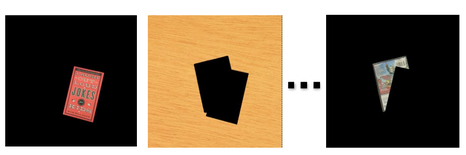
説明文の表現を調整するための質問文
“Please revise the description in color, text and illustration order.”

テキストプロンプトDB

推論：遮蔽時のセグメンテーション

In：入力画像
（物体遮蔽あり）

S4. テキストプロンプトDBに登録しているテキストプロンプトすべてを用いてセグメンテーション



出力された領域画像

S5. 出力された領域画像と商品画像で特徴点マッチングを行い正解領域画像を選択

Out：各商品の領域画像

実験

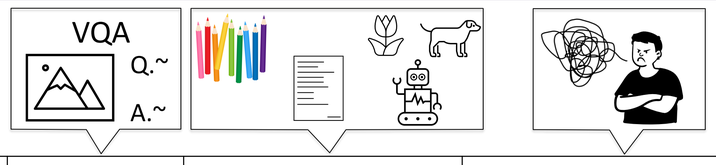
条件

● 画像データセット
- 平面物体(本,DVD) 10種類
- 机の背景画像 3種類

● 比較手法

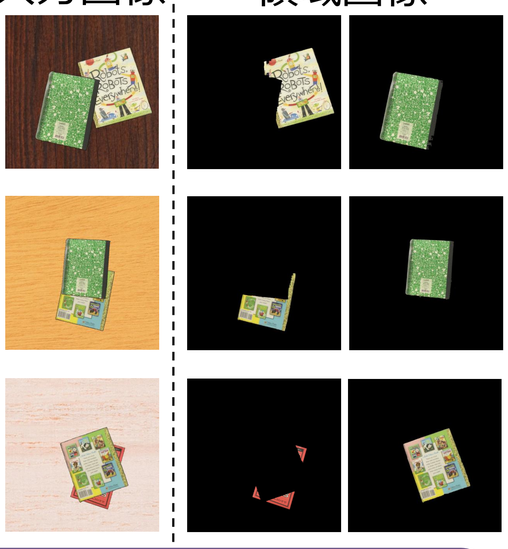
[Liang+,2018]
C_1: DeepLab v3+に物体遮蔽の訓練サンプルを人手で付与
C_2: EVF-SAMに単純なテキストプロンプトを付与
C_3: EVF-SAMに商品デザイン知識を基にテキストプロンプトを人手で付与
C_4: EVF-SAMにVQAのみからテキストプロンプトを自動で付与
O: 提案手法

結果



手法	視覚言語モデル	商品デザイン知識	手間		精度(mIoU)↑
			試行錯誤	マスク教師	
比較手法 C_1			**なし**	あり	0.92±0.04
比較手法 C_2			**なし**	**なし**	0.46±0.02
比較手法 C_3		✓	あり	**なし**	0.82±0.11
比較手法 C_4	✓		**なし**	**なし**	0.79±0.10
提案手法 O	✓	✓	**なし**	**なし**	0.83±0.10

入力画像　領域画像



テキストプロンプトデータベースを自動作成する事前準備によって試行錯誤の手間を省きつつ物体遮蔽が生じた場合の精度を改善