

AI を活用した入院患者の食事摂取量自動計測システムの開発

研究代表者： 青木 未来（医学系部門・助教）
共同研究者： 長谷川 達人（工学系部門・准教授）、四谷 淳子（医学系部門・教授）、
堀内 敦遥（工学部・学部生）、和田 海成（医学系研究科・大学院生）

概 要	
医療現場では、患者の摂取した栄養を把握する必要があり、病院内で提供された食事の残渣量の記録がされている。残渣量の記録は、看護師の目視で行われているため、時間がかかり看護師の負担となる。食事の残渣量記録を自動化することで、正確で看護師負担の少ない記録ができる。本研究では、基盤モデルを用いた Zero-shot 食事領域検出による、看護業務における残渣量の自動記録システムの開発を目的とした。食前食後の画像のペアを用いることで、先行研究では行われていなかった残渣量の推定を実現した。また、面積ではなく深度方向に減少する食事に対応するため、食器に対する食事の割合を用いた食事量推定手法を提案したが、結果として面積のみでの推定時から誤差が大きくなった。画像情報以外の情報を取り入れる必要があると考えられる。	
関連キーワード	食事残渣量推定、画像認識、深層学習モデル、栄養管理、看護業務

研究の背景および目的

医療現場では、患者の摂取した栄養を把握する必要があり、病院内で提供された食事の残渣量の記録がされている。残渣量の記録は、看護師の目視で行われているため、時間がかかり看護師の負担となる。食事の残渣量記録を自動化することで、正確な食事摂取量が把握でき、看護師の業務負担が軽減するため、患者の治療や健康維持に貢献できる。

既存研究では、食事画像から領域を切り出し、カロリー量を推定する研究や、センサベースで食事の摂取量を推定する研究がある。しかし、食前の画像のみを用いる研究は完食することを前提としており、食前食後の画像を用いた食事の残渣量推定

は行われていない。また、入院患者の負担やプライバシーを考慮するとセンサベースでの摂取量推定は現実的ではない。食事を認識するにあたって、既存研究ではまず皿領域を検出し、その内部の食事領域を k-means 法で検出している。この手法では皿が置かれている環境、皿と食事の色によって検出精度が低くなってしまうことが考えられる。そこで今回は、物体検出モデルとセグメンテーションモデルを組み合わせ、より正確な食事領域検出を試みることにした。

本研究では、基盤モデルを用いた Zero-shot 食事領域検出による、看護業務における残渣量の自動記録システムの開発を目的とした。

研究の内容および成果

1. 提案手法

本研究では、基盤モデルを用いた Zero-shot 食事領域検出により、看護業務における残渣量の自動記録システムを開発した。食事領域の検出には基盤モデルである Grounding DINO と、Segment Anything Model (SAM) の拡張手法である Segment Anything in High Quality (HQ-SAM) を用いた。Grounding DINO は物体検出モデル DINO を Vision-Language モデルに拡張した Zero-shot で運用可能な Open-set 物体検出モデルである。画像とテキストプロンプトを与えると、対

象となる物体の位置をバウンディングボックス (bbox) で返す。SAM はセグメンテーションのための基盤モデルである。1100 万枚の画像と、10 億以上のマスキラベルで構成された大規模セグメンテーションデータセットで訓練されており、高精度なインスタンスセグメンテーションを実現している。HQ-SAM は HQ-Output Token の導入により、元のトークンでは見落とされる境界領域や薄い構造領域に対応した SAM の拡張手法である。

我々の提案手法は、Grounding DINO を用いて bbox で該当領域を取得し、HQ-SAM を用いて

bbox 内の物体のセグメンテーションを行うことで残渣量の認識を行う。処理の流れを図 1 に示す。

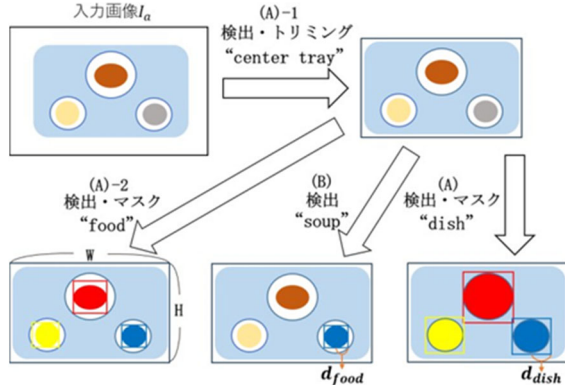


図 1 提案手法の流れ

(A)Grounding DINO+SAM-HQ による残渣料推定

1. Grounding DINO のプロンプトを”center tray”として、トレイ領域の候補となる bbox を検出し、検出された bbox の領域で元画像をトリミングする。
2. トリミングされた画像に対し、Grounding DINO のプロンプトを”food”として、食事領域の候補となる bbox を検出し、bbox を基準とした HQ-SAM の推定により食事領域のマスクを推定する。
3. 推定された食事領域マスクのピクセル数を合計し、トリミングした画像の面積で除算することで、式(1)により画像に対する食事領域の割合 f_{food} を求める。

$$f_{food} = \frac{P}{(W_a \times H_a)} \quad (1)$$

ここで P はマスクのピクセル数、 W_a はトリミング後の画像の幅、 H_a はトリミング後の画像の高さである。同様の作業を Grounding DINO のプロンプトを”dish”にして行い、どの食事がどの皿に入っているかの情報を記録した。

(B)soup 領域に対する例外対応

”soup”として検出された食事に対しては、食器に対する食事の割合 f_{soup} を推定した。これにより、

深さ方向に減少する食事の残渣量推定を行った。

(C)dish 領域に対する例外対応

食事の画像において食事領域の面積が皿領域の面積に近い場合は残渣量を 0 として扱った。これによって食後画像において皿全体を食事領域として検出してしまふケースを除外した。

2. 実験結果

本研究では、特別養護老人ホームで撮影した食前食後の画像 54 組を用いた。画像を撮影する際にカメラの高さや角度は固定せず、画像内には対象とする食事以外のトレイが含まれている場合もある。正解ラベルは看護師の目測で決定され、すべて残っている状態を 0、完食した状態を 10 とした 11 段階で付与された。目測で付与された正解ラベルとの差を図 2 に示す。誤差の絶対値の平均は 2.7 となった。一方、面積のみでの推定の場合は誤差の絶対値の平均は 2.6 であり、精度が悪化している。この原因として、食事前後で皿の状態が変わってしまうことが考えられる。

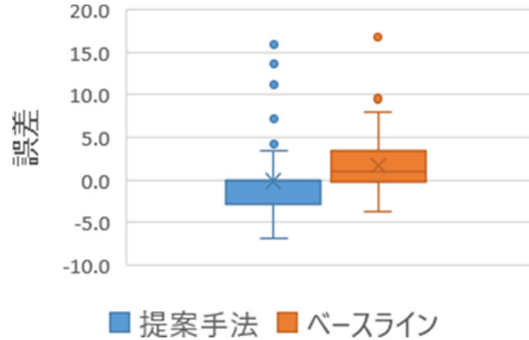


図 2 実験結果

3. まとめ

本研究では、基盤モデルを用いて食前食後の画像のみから食事の残渣量を推定する手法を提案した。本手法を用いることにより、Zero-shot 食事領域検出による食事の残渣率推定を行うことができた。また、食事領域の面積だけでなく、食器に対する食事の割合を用いた推定を行おうとしたが、面積のみでの推定よりも誤差が大きくなった。画像の情報のみで食事量を推定するには限界があり、他の情報を取り入れる必要があると考えられる。

本助成による主な発表論文等、特記事項および競争的資金・研究助成への申請・獲得状況

「主な発表論文等」

堀内敦遥, 中野雄太, 益本英明, 長谷川達人, 和田海成, 四谷淳子. "基盤モデルを用いた食事の摂取量推定", 情報処理学会第 86 回全国大会. 横浜, 2024.3.

堀内 敦遥, 益本 英明, 中野 雄太, 和田 海成, 長谷川 達人, 四谷 淳子, "基盤モデルを用いた食事の残渣量認識", 2023 年度電気・情報関係学会北

陸支部連合大会, オンライン開催, 2023.9.