

# 360度カメラの歪みに対応した物体認識モデルの構築

椎野 丞ノ進<sup>1</sup> 筑後 光<sup>1</sup> Sarthak Pathak<sup>2</sup> 梅田 和昇<sup>3</sup>

<sup>1</sup> 中央大学大学院理工学研究科精密工学専攻

<sup>2</sup> 芝浦工業大学工学部

<sup>3</sup> 中央大学理工学部精密機械工学科

E-mail: shiino@sensor.mech.chuo-u.ac.jp

## 1. 序論

現代社会において、リモート管理の重要性は急速に高まっている。その背景には、高齢化や人口減少による労働力不足、業務効率化の必要性、さらに都市化やグローバル化、デジタル技術の進展といった要因が挙げられる。建設業界では、高齢化や若年労働力の不足が深刻化しており、従来の労働集約型の管理手法から、効率的かつ省力化された管理方法への移行が求められている。また、都市化の進む中で、建設現場の規模や複雑性が増大し、物理的な距離や時間の制約を超えて現場状況を把握できる技術が必要とされている。こうした背景の中、360度カメラを活用したリモート管理技術が注目されている。米国 OpenSpace 社は、現場を 360 度カメラで撮影した画像を使ってクラウド上で施工管理を行うツール「OpenSpace」を開発した[1]。このシステムでは、ヘルメットやドローンに装着した 360 度カメラで現場を巡回し、毎秒 2 枚のペースで 360 度画像を撮影する。撮影された画像は、自動的に図面にマッピングされ、Google ストリートビューのように閲覧することが可能である。ストリートビュー画像は様々な分野で新たな需要が生まれており、近年ではリモートセンシング技術の一部として、変化検出への応用も進んでいる[2]。これにより遠隔からでも工事の進捗状況や変更箇所を効率的に確認することが可能となり、現場管理の大幅な効率化が実現されている。しかし、進捗状況が把握されていても現場内の工具、部品や設備などの特定の物体の個数や状態を詳細に把握・認識することは困難である。こうしたプロセスには依然として人の介入が必要であり、画像を人が手動でチェックし、時間的にも労力的にも大きな負担となり、効率性に欠ける。こうした課題を解決するために、自然言語を用いた物体検出を導入すれば、リモート管理の可能性は大幅

に拡大すると考えられる。具体的には、画像内の物体の個数や状態を検出し、自然言語による検索を可能にすることで、必要な情報を即座に取得できるようにする。一方で、自然言語を用いた物体検出を実現するためには、画像内のすべての物体を正確に認識する必要がある。しかし、360 度画像を平面展開する際に生成される正距円筒画像特有の歪みは、物体認識手法の精度を低下させる大きな要因になる。この問題を解決するためには、歪みを考慮した物体認識技術の開発が必要不可欠である。

本研究は、360 度画像を活用した自然言語検索が可能なシステムの基盤技術として、360 度画像の歪みに対応した物体認識の精度向上を目的とする。

## 2. 提案手法

本研究では、物体検出に YOLOv10[3]を用いる。また正距円筒画像に対応したセグメンテーションを行うために SAM2[4]を用いる。これらを用いることで歪みに対応した物体検出モデルを構築する。

### 2.1 セグメンテーションによる重心の検出

SAM2 は、特定のカテゴリに依存せず、任意の物体を高精度に検出・セグメント化することができる。特定の対象のみを検出する YOLO とは異なり、画像全体を対象に領域をセグメント化しているため、正距円筒画像にも対応することが可能である。さらに、SAM2 のセグメンテーションは、後続処理として物体の重心を特定するのに有効である。セグメンテーションにより領域分割された画像を用いることで、各物体に対応するセグメントの重心の計算を効率的かつ正確に行うことが可能となる。Fig. 1 は、歪みにより形状が変化してしまった人物に対して SAM2 を用いてセグメンテーションし、人物に対応する領域の重心を算出したも

のを表す。

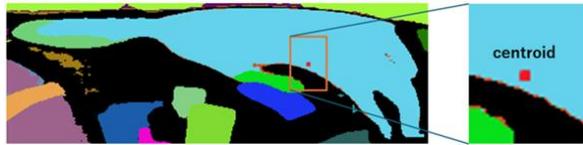


Fig.1 セグメンテーションによる重心の算出

## 2.2 重心を基にした画像補正

正距円筒画像は、一般的な画像と異なり、球面座標を平面上に展開したものであるため、画像を回転補正すると物体の形状が変わる。したがって、算出した重心を基準点として、対象物体が中心に位置するように画像を回転補正する。これにより、物体が歪みの影響を受けにくくなり、YOLOv10 による物体検出の精度向上が期待される。

## 2.3 歪み低減後の物体検出

回転補正後の画像に対して、YOLOv10 を用いて物体検出を行う。初期状態の正距円筒画像に YOLOv10 を適用した結果と回転補正後の画像に YOLOv10 を適用した結果を比較し、信頼度スコアを基に補正処理による歪み低減が物体検出精度に与える影響を評価する。

## 3. 実験

歪みによって形状が変化し、物体検出の精度が低下した対象に対して、画像補正することで精度が向上するかを検証する実験を行った。

本実験では、RICOH THETA Z1 を使用して撮影した 360 度画像を正距円筒画像に変換し、実験を行った。実験には物体検出アルゴリズム YOLOv10 および画像セグメンテーションアルゴリズム SAM2 を用いた。本研究は、提案手法の基礎的な検討を目的としており、工場環境ではなく、大学施設内で撮影した画像を使用して実験を実施した。

### 3.1 人を対象にした実験

#### 3.1.1 実験条件

撮影した正距円筒画像を Fig. 2 に示す。また、Fig. 3 に示される 6 名の人物(対象者 1~対象者 6)を検証対象とし、それぞれの物体検出精度を比較すること

で、提案手法の有効性を評価した。

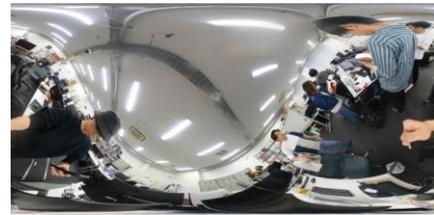


Fig. 2 補正前の正距円筒画像

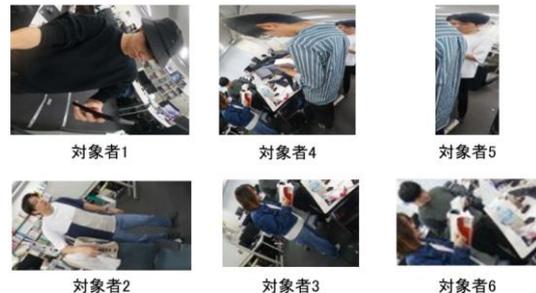


Fig. 3 検証対象の人物 6 名

#### 3.1.2 実験結果

実験結果を Fig. 4~Fig. 7 と Table 1 に示す。Fig. 4 は Fig. 2 の正距円筒画像に YOLOv10 を適用したものを表す。Fig. 5 は画像全体に対して SAM2 を用いてセグメンテーションを実行したものを表す。Fig. 6 は算出した重心を基に画像を回転補正し、物体が中心に位置するように調整したセグメンテーション画像を表す。Fig. 7 は補正後の正距円筒画像に YOLOv10 を適用した結果を表す。Table 1 は同様の手法で各対象人物に対して YOLOv10 を適用した結果とその信頼度を表す。「-」は対象が未検出であったことを示す。

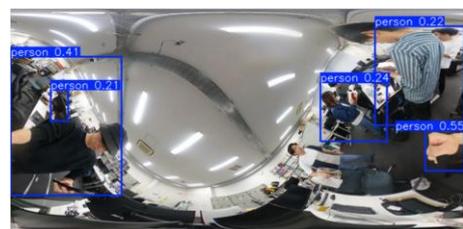


Fig. 4 YOLOv10 による検出結果(補正前)



Fig. 5 SAM2 によるセグメンテーション結果(補正前)



Fig. 6 SAM2によるセグメンテーション結果(補正後)



Fig. 7 YOLOv10による検出結果(補正後)

Table1 各対象人物とYOLOv10の信頼度スコア

	元画像	補正後
対象人物 1	0.41	0.91
対象人物 2	-	0.86
対象人物 3	0.24	0.77
対象人物 4	0.22	0.63
対象人物 5	-	0.41
対象人物 6	-	0.45
平均	0.29	0.67

## 3.2 トラックを対象にした実験

### 3.2.1 実験条件

撮影した正距円筒画像を Fig. 8 に示す. Fig. 9 に示すトラックを検証対象とし, 画像の歪みによる影響を低減した後の検出精度を測定し, 正確性の向上を確認した.



Fig. 8 補正前の正距円筒画像



Fig. 9 検証対象のトラック

### 3.3.2 実験結果

実験結果を Fig. 10~Fig. 13 と Table 2 に示す. Fig. 10 は Fig. 8 の正距円筒画像に YOLOv10 を適用したものを表す. Fig. 11 は画像全体に対して SAM2 を用いてセグメンテーションを実行したものを表す. Fig. 12 は算出した重心を基に画像を回転補正し, 物体が画像の中心に位置するように調整したセグメンテーション画像を表す. Fig. 13 は補正後の正距円筒画像に YOLOv10 を適用した結果を表す. Table 2 は補正前と補正後のトラックに対する YOLOv10 の信頼度を表す. 「-」は対象が未検出であったことを示す.

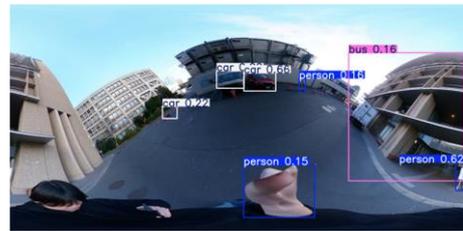


Fig. 10 YOLOv10による検出結果(補正前)



Fig. 11 SAM2によるセグメンテーション結果(補正前)

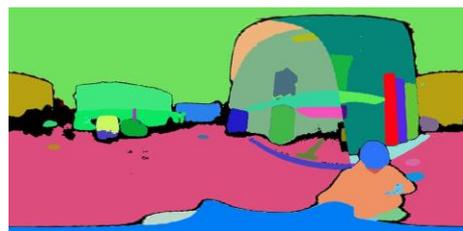


Fig. 12 SAM2によるセグメンテーション結果(補正後)

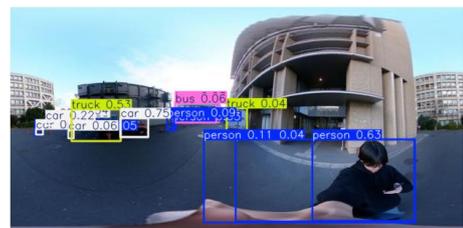


Fig. 13 YOLOv10による検出結果(補正後)

Table 2トラックとYOLOv10の信頼度スコア

	元画像	補正後
トラック	-	0.04

### 3.3 自転車を対象にした実験

#### 3.3.1 実験条件

撮影した正距円筒画像を Fig. 14 に示す。また、Fig. 15 に示す 2 台の自転車(自転車 1 と自転車 2)を検証対象とし、画像の歪みによる影響を低減した後の検出精度を測定し、正確性の向上を確認した。



Fig. 14 補正前の正距円筒画像



(b) 自転車 1

(a) 自転車 2

Fig. 15 検証対象の自転車

#### 3.3.2 実験結果

実験結果を Fig. 16～Fig. 21 と Table 3 に示す。Fig. 16 は正距円筒画像に YOLOv10 を適用したものを表す。Fig. 17 は画像全体に対して SAM2 を用いてセグメンテーションを実行したものを表す。Fig. 18 は自転車 1 の重心を基に回転補正し、物体が画像の中心に位置するように調整したセグメンテーション画像を表す。Fig. 19 は自転車 1 の補正後の正距円筒画像に YOLOv10 を適用した結果を表す。自転車 2 に対しても同様に、Fig. 20, 21 がそれぞれ自転車 2 の重心を基に回転処理を行い、物体が画像の中心に位置するように調整したセグメンテーション画像と補正後の正距円筒画像に YOLOv10 を適用した結果を表す。Table 3 は、補正前と補正後における自転車 1 と自転車 2 に対する YOLOv10 の信頼度を

表す。「-」は対象が未検出であったことを示す。



Fig. 16 YOLOv10 による検出結果 (補正前)



Fig. 17 SAM2 によるセグメンテーション結果 (補正前)



Fig. 18 自転車 1: SAM2 によるセグメンテーション結果 (補正後)



Fig. 19 自転車 1: YOLOv10 による検出結果 (補正後)



Fig. 20 自転車 2: SAM2 によるセグメンテーション結果 (補正後)



Fig. 21 自転車 2: YOLOv10 による検出結果(補正後)

Table 3 自転車と YOLOv10 の信頼度スコア

	元画像	補正後
自転車 1	0.57(person)	0.40(person)
自転車 2	-	0.44

### 3.4 実験結果の考察

人物を対象とした実験では、物体の重心を画像の中心に配置することで、歪みの影響を低減し、検出精度が向上することが確認された。特に、対象者 1 では信頼度が 0.41 から 0.91 に、対象者 3 では 0.24 から 0.77 に改善され、複数の対象において同様の傾向が見られた。この結果は、本手法が有効であることを示している。一方、トラックを対象とした実験では、回転補正によって検出は可能になったものの、信頼度は 0.04 と低い結果に留まった。これは、トラックの背面部が視覚的特徴に乏しいことや、学習データの不足が要因と考えられる。自転車を対象とした実験では、2 台の自転車に対して異なる検出結果が得られた。1 台は「person」と誤検出され、もう 1 台は「bicycle」として正しく検出された。この違いは、物体の角度や向きによる影響が大きく、側面が明確な場合は正しく認識される一方、後方からの視点では誤認識が発生しやすいことが示唆される。よって、物体の外観や向き、学習データの多様性が検出精度に与える影響が大きいことが明らかになった。また自転車 2 に示す結果に関して、2 台の自転車が 1 つのバウンディングボックスにまとめられていることが確認された。この要因として、YOLO のバージョンによるアルゴリズムの違いが影響していると考えられる。YOLOv8 や最新の YOLO11 のバージョンでは、NMS (Non-Maximum Suppression) を適用することで、同じ物体に対して複数のバウンディングボックスが生成された場合に、IoU (Intersection over Union) の閾値に基づき、最も高い信頼度の高いバウンディングボックスを残し、他を削除する処理が行われる。そのため、近接した物

体であっても個別に検出されやすい。一方、YOLOv10 では NMS が廃止され、代わりに one-to-one matching が採用されている。この手法では、各物体に対して 1 つのバウンディングボックスのみを割り当てるため、物体が密集している場合に、複数の物体が 1 つのバウンディングボックスとして検出されたと考えられる。総じて、Fig. 22 に示すとおり、密接した物体を個別に識別する必要がある場合、NMS を適用している YOLOv8 や YOLO11 の方が適していると考えられる。加えて YOLO11 は本研究で用いた YOLOv10 と比較してパフォーマンスが高い[5]とされており、今後は YOLO11 の利用を検討していく。



Fig. 22 自転車 2: YOLO11 による検出結果

## 4. 結論と今後の展望

本研究では、360 度画像における歪みに対応した物体検出の新たな手法を提案し、提案手法の有効性を検証するために、人物、トラック、自転車を対象にした検出実験を行った。提案手法では、まず 360 度画像を正距円筒画像に変換し、初期状態で YOLOv10 を適用して物体を検出する。次に、SAM2 を用いてセグメンテーションを行い、物体の重心を算出する。その後、正距円筒画像を回転補正することで歪みを低減した状態で再度 YOLOv10 を適用する。これにより、初期の検出結果と回転補正後の検出結果を比較し、検出精度の向上を比較した。実験の結果、提案手法は 360 度画像の歪みに対応する有効なアプローチであることが示された。一方で、物体の外観、角度、向き、および訓練データの多様性が検出精度に大きく影響することも明らかとなった。また近接した物体が 1 つのバウンディングボックスとして認識されるケースも確認された。これは YOLO のバージョンによるアルゴリズムの違いに起因している可能性があり、検出対象の特性や目的に応じて、どの手法を採用するか慎重に検討する必要があることが分かった。

本研究は自然言語を用いた物体検出や異常検知を実現するための基盤技術として考えられる。今後は、

CLIP[6]などの自然言語モデルを取り入れることで、自然言語を用いた物体検出の実現を目指す。また、本研究で物体の外観、角度、向き、および訓練データの多様性が物体検出性能に大きく影響することが明らかになった。これを踏まえ、建設現場等、用いる環境に適用するため、現場特有の環境に対応したデータセットの構築を検討する。最終的には建設現場の安全管理や作業効率向上に貢献できる物体検出システムの構築を目指す。

## 参考文献

- [1] <https://www.openspace.ai/ja/products/capture/>
- [2] W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *\*Remote Sensing\**, vol. 12, no. 10, p. 1688, (2020)
- [3] Wang, Ao, et al. "Yolov10: Real-time end-to-end object detection." arXiv preprint arXiv:2405.14458 (2024).
- [4] Ravi, Nikhila, et al. "Sam 2: Segment anything in images and videos." arXiv preprint arXiv:2408.00714 (2024).
- [5] <https://docs.ultralytics.com/ja/models/>
- [6] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, (2021).