

摘要

近年の建築では BIM (Building Information Modeling) と呼ばれるコンピュータ上に現実と同じ建物の立体モデルを再現し , 可視化するワークフローが注目されている . 従来の配管 BIM は高精度な Lidar センサを用いて配管モデルの推定を行なわれていたが , 振動に弱く高価である . そのため , Lidar センサより安価である RGBD カメラを採用する . 本研究は従来の点群データのみを用いた 3D 再構築を行わず , 取得画像と関連する点群データに基づき配管の 6D 姿勢推定を行い 3D 再構築の高精度化を目的とする . また , 姿勢推定の従来の方法では手作業で行われていたが , 深層学習を取り入れることにより高精度かつ高速化を図る .

謝辞

本稿の内容は，著者が京都大学大学院工学研究科機械工学専攻メカトロニクス研究室（吉川研）での博士後期課程において，そして2007年度現在，講師として在籍している立命館大学理工学部ロボティクス学科において学んだ内容をまとめたものです。あまり明文化されない，しかし非常に重要なこのような知識を与えて頂いた，吉川研と立命館大学ロボティクス学科のすべての方々に，深く感謝の意を表します。

なお，本稿の表紙には，指導教員として川村貞夫教授の名前がありますが，これは修士論文，卒業論文のテンプレートとして出力されたものです。したがって，本稿の文責は川村教授ではなく，すべて著者（金岡）にあることをここに明記しておきます。

目次

第 1 章 序論	1
1.1 研究背景	1
1.2 既存研究	3
1.3 研究目的	5
1.4 本手引の構成	5
第 2 章 深層学習による配管 6D 姿勢推定	7
2.1 配管 6D 姿勢推定方法	7
2.2 配管の特徴を活かしたアイソメ図作成方法	8
2.3 ネットワーク構造	8
2.3.1 全体構成	8
2.3.2 RXD ネットワーク	8
第 3 章 配管データセット	11
3.1 使用機材	11
3.2 物体検出のデータセット収集	12
3.3 6D 姿勢推定のデータセット収集	13
第 4 章 実験	15
4.1 評価指標	15
4.2 結果と考察	16
第 5 章 結言	21
参考文献	23
付録 A 数式の記述	25
A.1 記述例	25

図 目 次

1.1 アイソメ図の例	2
1.2 従来のアイソメ図取得方法	2
1.3 YOLO モデルの検出の流れ	3
1.4 Gen6D ネットワーク構造	5
2.1 RGB-D カメラを用いた深層学習によるアイソメ図作成方法	7
2.2 配管の検出例	8
2.3 RXD ネットワーク構造図	9
2.4 Sigmoid 関数のグラフ	10
2.5 RXD ネットワーク構造図	10
3.1 暗闇での RGB-D カメラの撮影	11
3.2 Intel Realsense L515	12
3.3 Colmap を用いた曲管の点群データ	13
4.1 Intersection over Union(IOU)	15
4.2 適合率と再現率	16
4.3 適合率と再現率	16
4.4 適合率と再現率	17
4.5 適合率と再現率	17
4.6 適合率と再現率	18
4.7 適合率と再現率	19

表 目 次

4.1 物体検出ネットワークの実行結果	16
4.2 テスト画像を用いた検出結果	18

第1章 序論

配管は気体、液体、粉粒対などの流体を輸送や配線の保護などを目的とする管のことである。配管は様々な場面で使用されており、電気配線やケーブルを保護する電気配管や、生活に必要な水を家庭や学校などに輸送する水道管などが挙げられ、私達の生活において重要な役割を担っている。そのため、配管を運用するにあたって常に耐久性と安全性を保ち続ける必要性がある。

1.1 研究背景

BIM とは、Building Information Modeling の略称で、建築物や土木構造物などの情報をコンピュータ上に現実と同じ建物の立体モデルを形成し、設計から維持管理までのプロセスをデジタル化する新しいワークフローの一環である。この BIM モデリングはこれまでの 3D モデリングとは大きく異なる。まず、従来の 3 次元モデリングでは 2 次元上で作成した図面を元に、3 次元の形状を形成し組み立てる手法であった。そのため、3D モデルに修正箇所があった際に、2 次元の図面を全て修正してから再度構築する必要があり大きな手間が生じていた。しかし、この BIM 手法は一つのデータを修正すると全てのデータが連動し、関係する図面の該当箇所が自動修正され、従来の 3D モデリングよりも高校率で作業を行うことが可能になる。

配管は建築物の中でも日常生活に欠かせない存在である。配管は生活に必要な物資を運用したり保護する役割があり、常に安全性と耐久性を満たす必要がある。その配管の図面を作成する際にはアイソメトリック（アイソメ）図と呼ばれる立体を斜めから見た視点で表示した等角図が用いられる。このアイソメ図の最大の特徴が配管のルートを人目でイメージしやすくなることだ。設計図には平面図や立体図、系統図など様々な種類の図面を使用するが、配管の場合、配管同士が複数にも重なり合っているため左右上下からの視点では見分けることが困難である。そのため、アイソメ図は図面を立体的に描画する手法を扱えるため、配管のルートや交差する配管の前後関係をイメージしやすくなる。

アイソメ図を取得するためにはこれまでに Light Detection and Ranging(LIDAR) センサーと呼ばれるレーザー光を使用して離れた場所にある物体の形状や距離を測定できるセンサーを使用していた。RGB 画像では距離情報を取得できないことから、3 次元モデリングは比較的困難とされているが LIDAR センサーは距離情報を取得できることから、オブジェクトの奥行きを点群データとして扱えるようになる。LIDAR センサーは 3 次元情報を取得できる点や測定範囲や精度が良いというメリットがあるが、その反面他のセンサーと比較すると高価であるというデメリットを抱

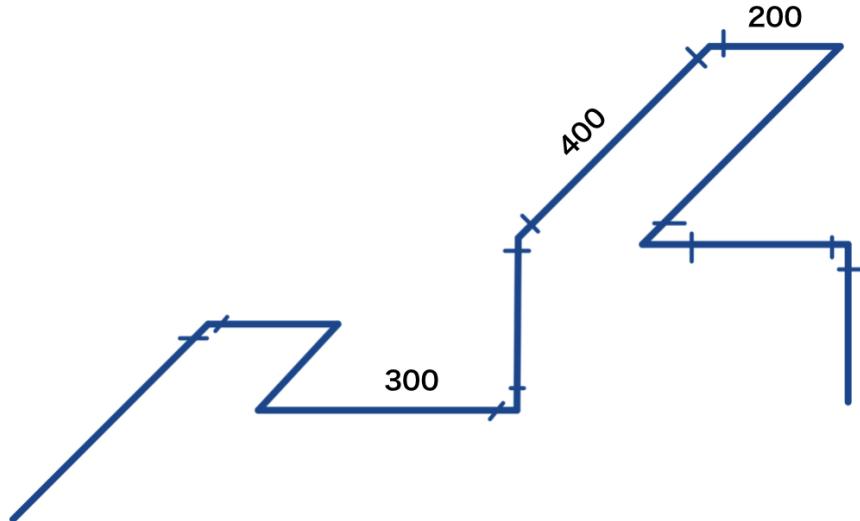


図 1.1: アイソメ図の例

えている。そのため、広く一般的に使用するためには安価なセンサーでデータ収集し図面を作成できることが望まれる。このような背景から近年RGBカメラやRGB-Dカメラを用いたLIDARセンサーよりも安価な機器を用いた認識手法が研究されている。その認識手法には近年、機械学習による物体検出と姿勢推定のタスクがコンピュータビジョンにおいて幅広く研究されている課題である。

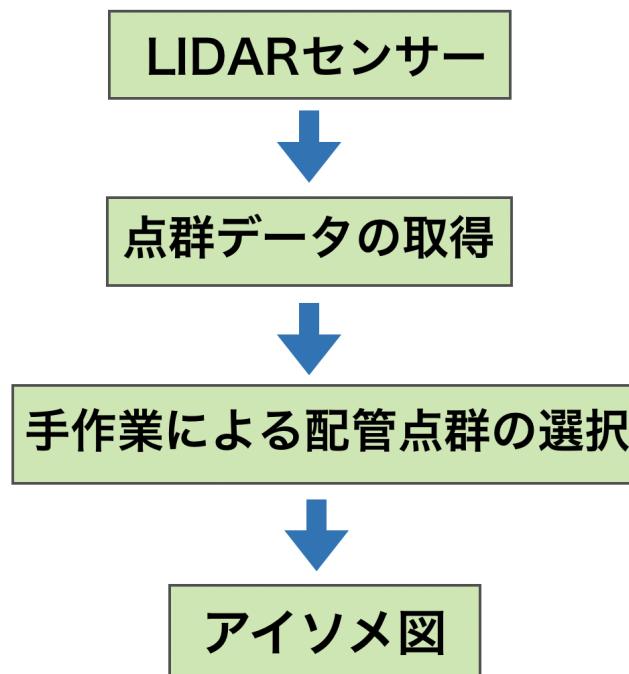


図 1.2: 従来のアイソメ図取得方法

1.2 既存研究

機械学習による画像認識分野は画像に写る物体を識別し位置を特定する物体検出や画像の画素ごとに識別を行うセグメンテーション、オブジェクトの位置情報に加えて向きを推定する姿勢推定問題など様々な分野で研究がなされている。物体検出は画像内で認識したいオブジェクトがどこに存在しているのかをバウンディングボックスを用いて検出するのが一般的である。その代表的なモデルとして YOLO を紹介する [10]。このモデルはほぼ同時期に発表された Fast R-CNN と同様に、物体検出に大きな影響を与えた [12]。Convolutional Neural Network(CNN) と呼ばれる畳み込みという操作を加えたニューラルネットワーク構造を使用してオブジェクトを検出する。CNN の中には畳み込み層やプーリング層といつたりいくつかの個性的な機能を備えた層が含まれ、人手による作業を必要とせず得られた特徴をもとに領域を予測することができる。YOLO の特徴は従来までは境界設定と物体検出を 2 段階に行っていた作業を一度に行うことで推定速度の高速化を行うことができた。図 1.2 に YOLO のネットワーク構造示す。

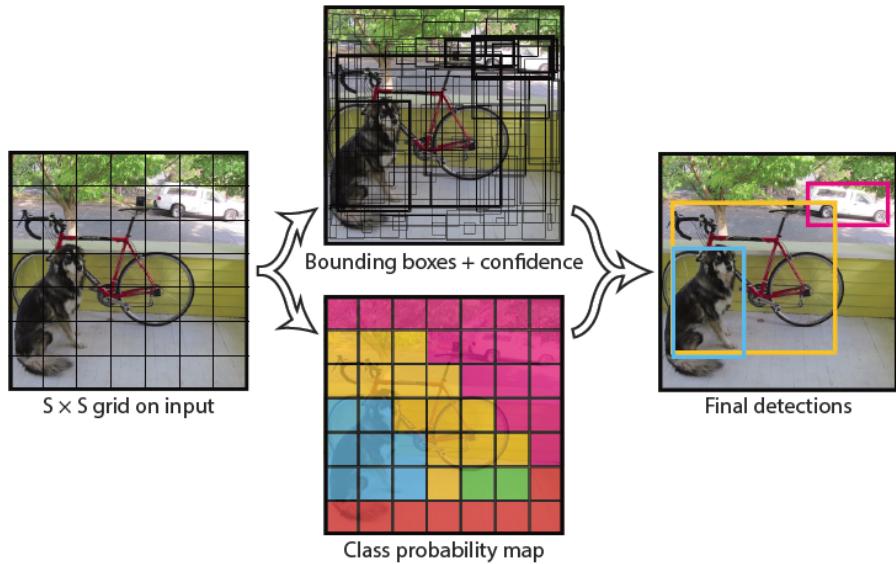


図 1.3: YOLO モデルの検出の流れ

まず、入力画像を $S \times S$ のグリッドセルに分割し、各グリッドセルで複数個のバウンディングボックスと各バウンディングボックスの信頼度を計算する。物体の中心がグリッドセルに存在していた場合に、そのセルが物体を検出するように学習する。次に、バウンディングボックスの推定では各グリッドセルに B 個のバウンディングボックスを持ち、それらのボックスの信頼スコアを予測する。信頼スコアとは背景ではなく物体が含まれている確率のことである。次に、各グリッドセルは複数のクラスに対する条件付き確率を予測する算出された条件付きクラス確率と一つ前の個々のバウンディングボックスの信頼スコアを掛け合わせることで、バウンディングボックス毎のクラスに対する信頼スコアを得ることができる。このスコアを使用しどのバウンディングボックスが正解の物体を推定しているのかを判断している。これ以後、End-to-End モデルと呼ばれる入力層から出力層まで全層の重みを一辺に学習す

る手法が物体検出の中で主流となった。

一般的な物体検出では RGB 画像を用いた手法が多いが、カラー画像に Depth 画像を取り入れた物体検出方法も存在する。Depth 画像は物体の奥行き情報や、外光の影響を受けづらいため暗闇の中でも安定してオブジェクトの特徴を捉えることができる点が優れている。RGB-D 画像を用いた物体検出はカラー画像と深度画像をそれぞれ両方畳み込みした値を最後の全結合層で結合するのみの手法が一般的であった [14]。しかし、この手法ではカラー画像と深度画像のそれぞれの特性を維持することはできず、最大限 RGB-D 画像の利点を活かすことができていなかった。そこで SPnet モデルではクロス強化モジュール (CIM) を提案することで RGB 画像と深度画像から抽出された特徴を維持したまま統合する機能を実現可能にした [13]。

次に、6D 姿勢推定問題について紹介する。姿勢推定問題では物体検出と同様に RGB カメラや RGB-D カメラを使用した推定方法がある。RGB カメラの姿勢推定問題では古典的な方法はキーポイントを検出し、既知のオブジェクトモデルを参照することによって推定する [1, 2, 3]。また、最近の研究では 2 次元上でキーポイントを予測し [5]、PnP によって姿勢を算出することが可能になる [4]。また、画像からオブジェクトの姿勢を直接推定する手法も提案されている [6]。RGB-D カメラを用いた姿勢推定問題では奥行き情報を使用できるため参照できる情報量の増加により精度が向上している [7]。また、Densefusion は同様に RGB-D 画像を用いて姿勢推定問題に取り組んだ [8]。姿勢検出においてオクルージョンと呼ばれる手前にある物体が後ろにある物体を隠す問題が課題となっていたが、独自のネットワーク検出方法により、他のネットワークよりも優れた精度を示している。6D 姿勢推定を行うネットワークである Generalizable Model-Free 6-DoF Object Pose Estimation from RGB Images(Gen6D) を紹介する [9]。姿勢推定に必要な主なデータセットは 3 次元データやカラー画像、深度データなどが代表的である。しかし、3 次元データをデータセットに使用するには事前に、認識したいオブジェクトの 3D モデルを作る必要があるため、大きな手間が生じてしまう。そのため、Gen6D はデータセットに 3D データを必要とせずカラー画像のみで物体の姿勢推定を行える手法を提案した。データセットには Colmap と呼ばれる 2 次元画像から 3 次元点群を再構築するために使用されるソフトウェアが用いられている [11]。Colmap に使用される 2D 画像はオブジェクトを異なる視点から撮影された画像を複数枚利用することで 3 次元情報を復元することができ、その点群データを学習して物体の 6D 姿勢を推定する。

Gen6D のネットワーク構成について図 3 に示す。まず、Detector と呼ばれる工程では参照画像の情報をもとに認識したいオブジェクトの領域を検出する。次の工程である、Selector では Detector で得られた領域の画像と最も近い視点を持つ参照画像を複数枚ある中から 1 つ抽出する。これは選択された参照画像の視点をテスト画像の視点とほぼ同様とみなし、誤差は生じますがオブジェクトのポーズの初期姿勢を形成する。最後の工程では先程得られた姿勢の改良を試みる。まず、参照画像から近い視点の画像をさらに 6 枚選択し、全参照画像間の平均と分散を算出し、初期に求められた姿勢の情報を改善して最終的な結果を予測する。この研究のメリットとして RGB 画像のみを用意することで物体の姿勢を推定できるため、データセットの作成は非常に容易である。しかし、この Gen6D をしようするにあたって問題点が 2 つある。まず、一つ目に RGB 画像は距離情報を持たないため、物体のスケール情

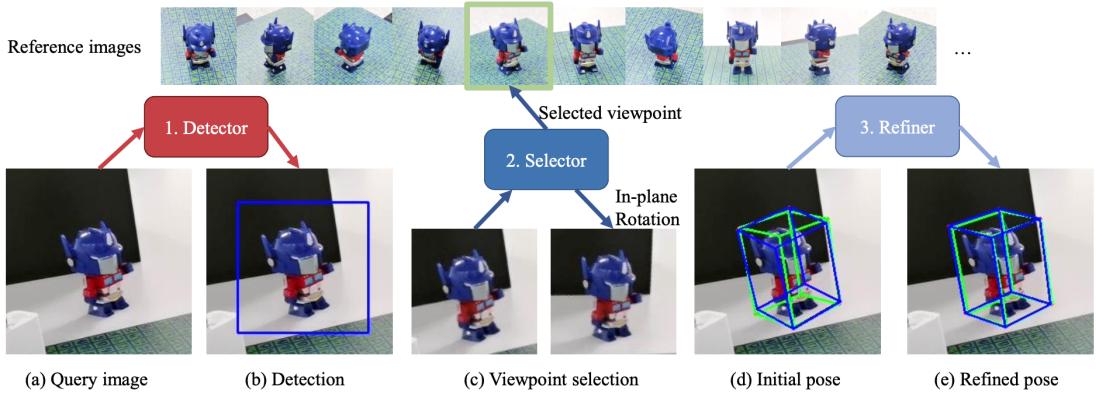


図 1.4: Gen6D ネットワーク構造

報を明示することはできない。2つ目は一度の推論で複数の物体の姿勢推定を行うことができない点である。

アイソメ図を作成するにあたって、配管同士の距離情報は必ず明示する必要がある。また、暗闇の状況下でも配管を認識可能にするために、本研究では RGB カメラではなく RGB-D カメラを使用し、Depth 画像を活用したネットワーク設計を試みる。

1.3 研究目的

本研究では RGBD データを使用した深層学習による配管 6D 姿勢推定を行い、RGBD カメラを用いることによる安価な機器での姿勢推定の実現を試みる。また、既存の RGB 画像のネットワークに Depth 画像を組み込んだモデルを提案し、認識精度向上と推定速度の高速化を目標とする。本研究の貢献は以下のようになる。まず一つ目は深層学習による RGB 画像と Depth 画像を用いた物体検出ネットワーク (RXD) の提案である。RGB 画像と Depth 画像からそれぞれ抽出された特徴を結合する RxDLayer を導入し、他のネットワークと比較し RXD ネットワークの有効性を示した。

2 つ目は既存の 6D 姿勢推定ネットワーク (Gen6D) の複数物体検出を可能にさせたことである。配管は単体ではなく複数の管が張り巡らされているため、複数の認識を可能にする必要がある。RXD ネットワークでは画像内部にある配管全てを網羅し、それぞれの物体の中心ピクセル座標とスケールを推定することができる。

3 つ目は本研究の最終目的であるアイソメ図を作成するにあたっての必要不可欠な配管距離測定である。アイソメ図は配管の向きだけでなく、距離情報を図面に示す必要がある。そのため、Depth 画像を用いることでネットワークによって認識された情報をもとに、配管の距離情報を算出することを可能にした。

1.4 本手引の構成

本論文の構成は以下のようになる。第一章では研究背景、既存研究、研究目的について述べる。研究背景では、Building Information Modeling(BIM) についてや從

来のアイソメ図の取得方法について述べる。既存研究では、6D 姿勢推定と物体検出のそれぞれのネットワークを紹介する。研究目的では、本研究の目的及び貢献について述べる。

第2章では、データ収集から配管アイソメ図までの方法や流れについて説明する。また、RXD ネットワークの提案と構造図について紹介する。第3章では、データセットの概要について述べる。データセットを収集する機器についてや RGB-D カメラを使用するに適した配管のデータセットについて紹介する。第4章では、物体検出や姿勢推定をテスト画像の結果や評価指標に基づいた数値より考察する。第5章は結論である。

第2章 深層学習による配管6D姿勢推定

従来のアイソメ図取得には LIDAR センサーにより 3 次元点群を取得し図面を作成していたが、センサーが高価であるというデメリットを抱えていた。そのため、本研究では LIDAR センサーよりも比較的安価な RGB-D カメラを用いてデータセット収集から深層学習やアイソメ図作成までの流れを紹介する。第 2.1 節では RGB-D カメラを用いた深層学習による配管 6D 姿勢推定の手順を述べる。第 2.2 節では物体認識のネットワーク設計を詳しく説明する。

2.1 配管 6D 姿勢推定方法

RGB-D カメラを用いた深層学習による配管のアイソメ図作成の手順はデータ収集、物体検出、6D 姿勢推定、アイソメ図作成の 4 つに分けられる。図 2.1 にそのプロセスの流れを示す。

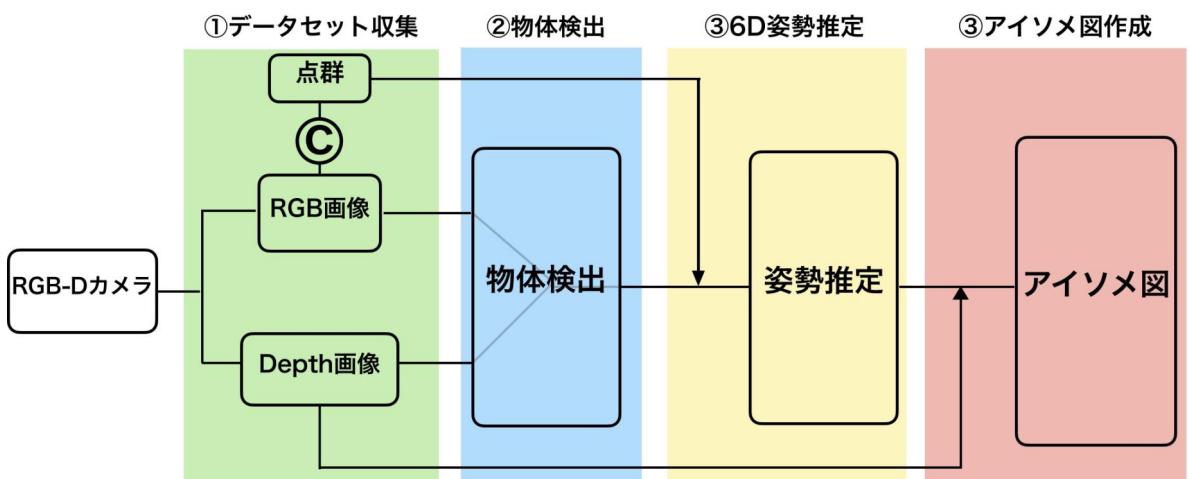


図 2.1: RGB-D カメラを用いた深層学習によるアイソメ図作成方法

まず、RGB-D カメラを用いて RGB 画像と Depth 画像を取得する。これらの画像を使用して物体検出ネットワークより複数の配管検出を行う。次に、物体検出で認識された配管の姿勢を推定する。その際に、データセットとして RGB 画像から 3 次元復元ツールである Colmap を使用して点群データを取得する。この Colmap は Structure from Motion(SfM) という技術で異なる視点からの写真を使用して 3 次元形状を復元する写真解析ソフトウェアである。Colmap から得られた点群データをもとに姿勢推定問題に取り組む。最後に姿勢推定された結果を用いてアイソメ図を

作成するが、図面には配管の距離情報を示す必要がある。そのため、Depth 画像を用いることで認識された配管のスケールを算出することができる。以上のステップを踏むことで RGB-D カメラからアイソメ図を作成することができる。

2.2 配管の特徴を活かしたアイソメ図作成方法

配管の形状には配管固有の特徴が存在する。図 2.2 に一部配管の例を示す。配管は両端部分を除くと直線であるという特徴があるため、両端の曲管又は T 字管がどの方向を向いているのかを認識できればその間を直線で結ぶことでアイソメ図を作成することができる。そのため、本研究においては配管全体を認識するのではなく、画像内の曲管及び T 字管を検出し姿勢を推定する。それに加え、両端の曲管又は T 字管間の距離を Depth 画像を用いて算出することで図面を作成することができる。

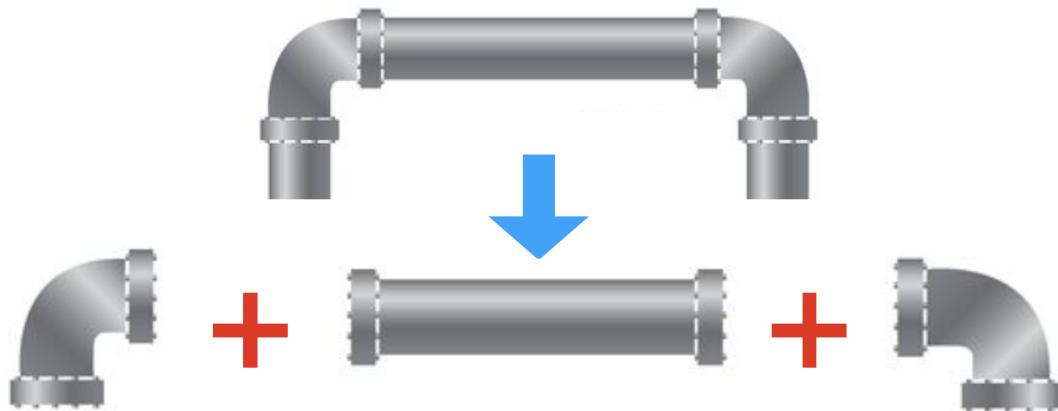


図 2.2: 配管の検出例

2.3 ネットワーク構造

2.3.1 全体構成

本研究では曲管及び、T 字管の物体検出と 6D 姿勢推定を深層学習を用いて推定する。その全体構成を図 2.3 に示す。物体検出ネットワークでは RXD Layer 用いた RXD ネットワークを提案する。RGB 画像と Depth 画像を RXD Layer に挿入することでそれぞれの画像から抽出された特徴を結合し認識精度向上を図ることができる。また、姿勢推定ネットワークでは Gen6D モデルを使用する。Gen6D は複数物体の検出ができないが、RXD ネットワークから認識された複数の配管のピクセル座標とスケールを Gen6D の Selector にそれぞれ渡すことで Refiner を通して最終的に画像内の全ての曲管及び T 字管の姿勢を求めることが可能になる。

2.3.2 RXD ネットワーク

物体検出に使用する RXD ネットワークについて紹介する。RXD ネットワークの構成を図 2.4 に示す。RGB カメラから取得された RGB 画像と Depth 画像をそれぞ

れ Convolutional set に挿入することで特徴を抽出することができる。深層学習は基本的に層を深くすることで認識精度が向上するため、RXD ネットワークでは複数回畳み込み層を使用している。また、畳み込みの際には Batch Normalization(BN)、ReLU、Max Pooling(MP) を各層に取り入れている。Batch Normalization は各バッチのデータを使用し正規化を行う。その結果、出力が適度に分散され、勾配消失などの問題が起こりにくくなり、学習が適切に進む。特に深めのネットワークを使用したときに、数カ所に挟むことで効果を得る。次に Max Pooling とは CNN で用いられる基本的なプーリング層である。最大値プーリングではカーネル内の最大値のみを残すプーリング処理である。これらの層を複数利用することで特徴をより濃くすることができる。

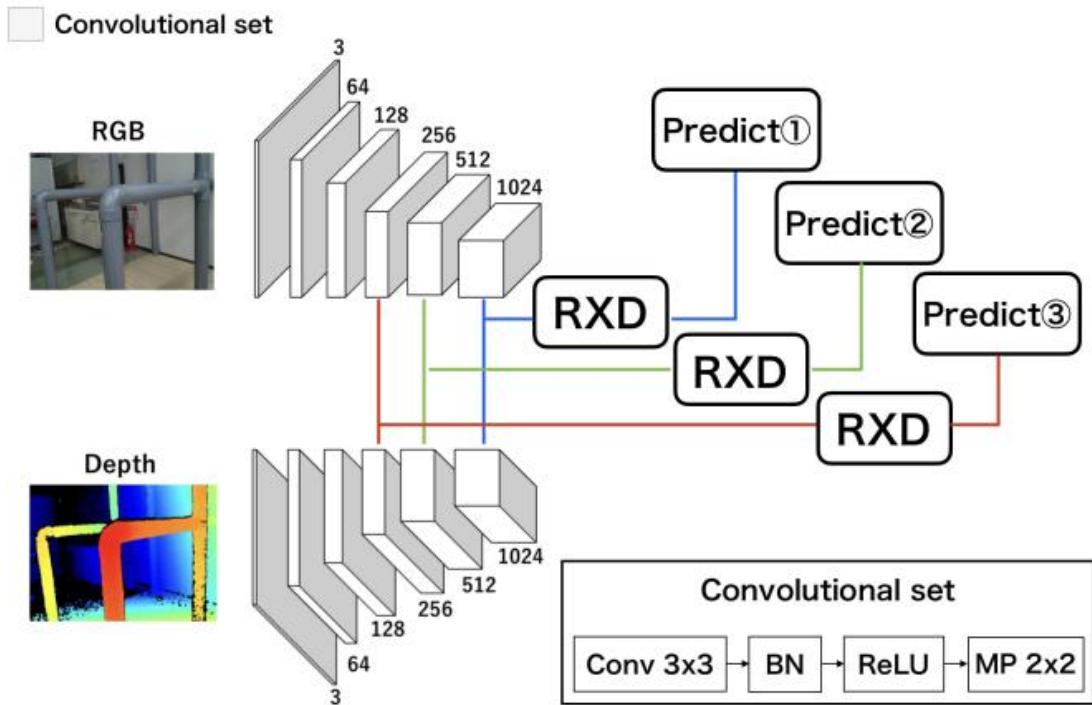


図 2.3: RXD ネットワーク構造図

RXD 層の内部構成を図 2.5 に示す。RXD 層の中身では RGB 画像と Depth 画像が RXD ネットワークで畳み込まれたデータを結合する役割を担っている。まず、RGB 画像と Depth 画像をそれぞれ畳み込み特徴マップを取り出す。それらのデータを Concatenate 関数を用いて連結させる。次に、結合されたデータを畳み込んだあとは Sigmoid 関数という活性化関数を使用する。通常の活性化関数には ReLU 関数が使用され、入力が 0 以下の時は 0 を、0 より大きい時はその値を出力する関数である。Sigmoid 関数は ReLU 関数とは違い、入力値 x の値に依らず、0 ~ 1 の数値に変換して出力する。次に、Sigmoid 関数によって出力された値をそれもとの RXD ネットワークから入力されたデータと乗算する。このステップにより RGB 画像と Depth 画像の相関を利用し特徴マップの表現力を強化することができる。これによって得られたそれぞれの値を Concatenate 関数を用いることで結合し、2 度の畳み込み層を経ることで出力結果が求められる。

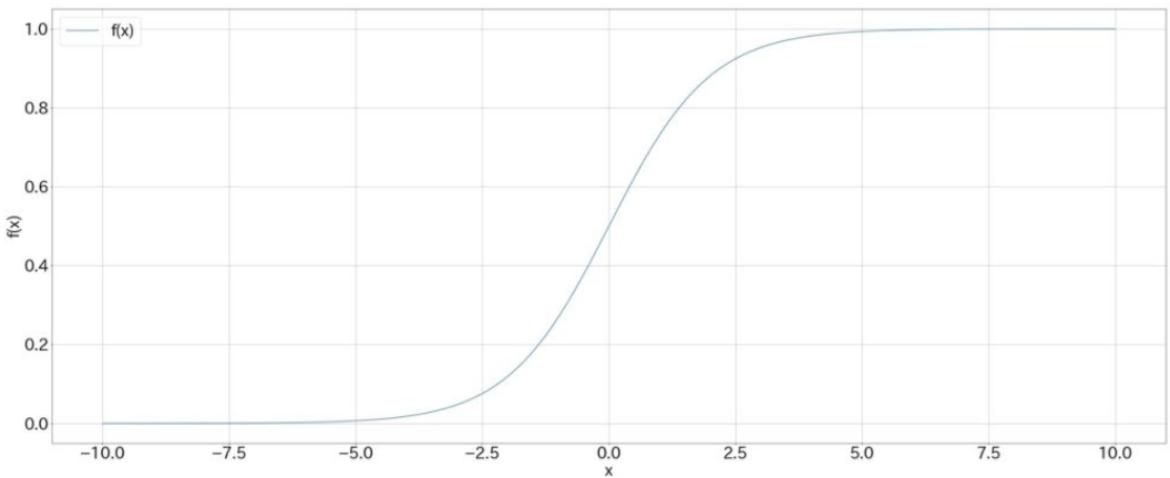


図 2.4: Sigmoid 関数のグラフ

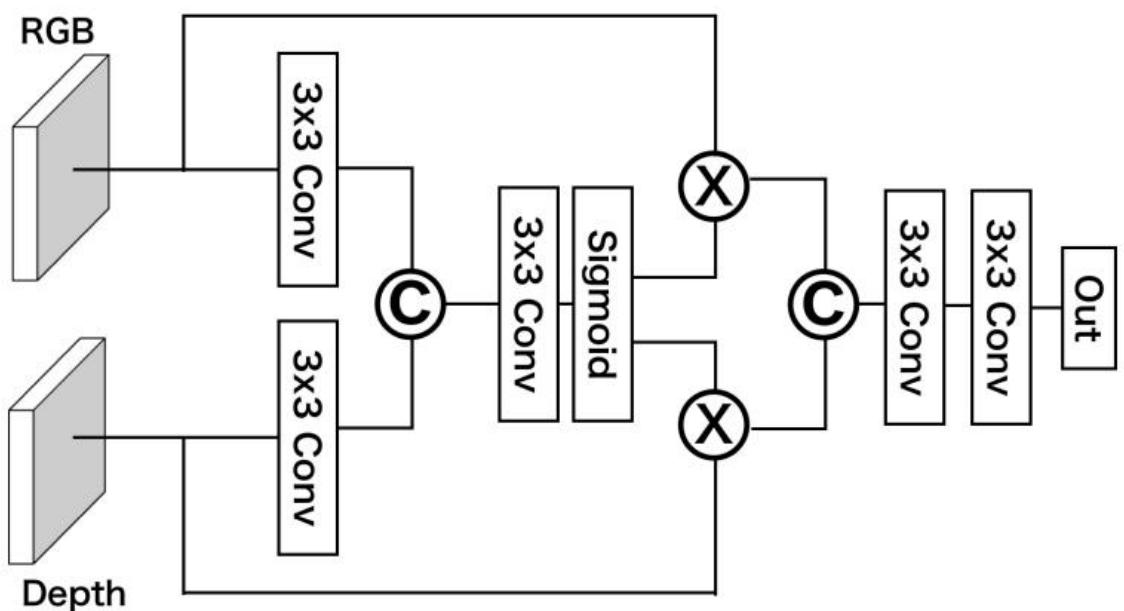


図 2.5: RXD ネットワーク構造図

第3章 配管データセット

前章では、アイソメ図を作成するためのネットワーク構造を提案した。本章では、深層学習に用いるデータセットの収集及び、作成方法について論じる。

3.1 使用機材

データセットの取得にはRGB-Dカメラを使用する。従来の方法ではLIDARセンサーを用いて配管の3Dデータやアイソメ図を作成していた。しかし、LIDARセンサーは高価であり、一般的に使用することが困難であるという欠点を抱えていた。そのため、RGB-DカメラはLIDARセンサーよりも比較的安価であるため本研究のデータセット収集に使用する。次に、RGBカメラではなくRGB-Dカメラを使用する利点を紹介する。RGB-DカメラのDepth画像にはたくさんのメリットが存在する。まず、1つ目にDepth画像には距離情報を取得できるという点である。配管のアイソメ図には配管のそれぞれの部位の長さを正確に示す必要がある。そのため、RGB画像には距離情報を含まれていないことからスケールを求めるにはDepth画像が重要になるのだ。2つ目に光の明暗に影響されない点である。RGB画像は撮影する環境が暗闇の場合、画像には何も映らない。これはRGB画像が光に反射された物体の度合いを数値化しているため、極端に明るすぎたり暗すぎるとRGB画像が活用できなくなる。特に配管が設置されている地盤地下や天井裏などの照明を当てることが困難な環境ではDepth画像が必要になる。3つ目に配管が背景色と同様の色を示していた場合に、区別が容易に可能であるという点である。RGB画像では色の違いが判断できないが、Depth画像は距離情報の違いを示すことができるため、背景と異なる物体として認識可能になる。以上の点より本研究にはRGB-Dカメラを採用した。

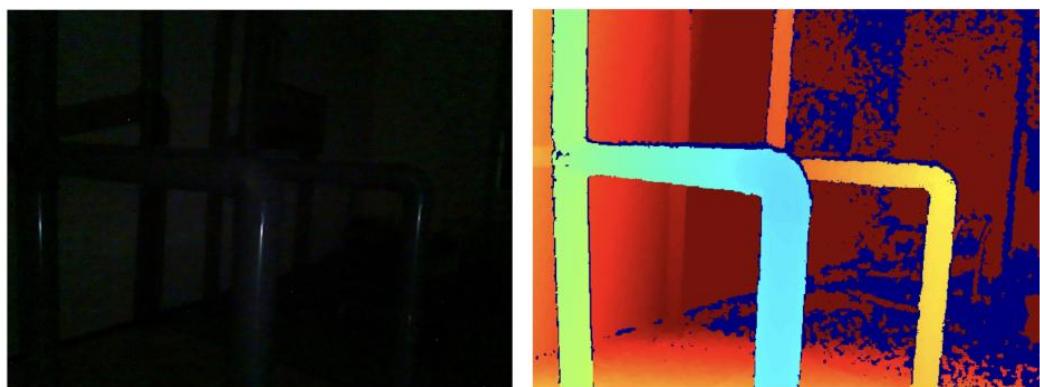


図 3.1: 暗闇でのRGB-Dカメラの撮影

カメラはインテル社製の Intel Realsense L515 を使用した。Realsense L515 を使用した理由は Realsense カメラの中でも屋内に適した RGB-D カメラであるからだ。このカメラは外光の影響を受けやすいが、屋内の環境であればその影響を受けないため、配管などの室内で多く使用される環境では非常に適していると判断した。



図 3.2: Intel Realsense L515

しかし、Realsense L515 は仕様上、RGB カメラと Depth カメラの位置が異なるため、撮影した際に両方の画像を比較すると画角に差異が生じてしまう。これは、データセットのラベリングを行う際に配管のピクセル座標にそれぞれの画像で異なると認識の精度に大きな誤差が生じてしまう。そのため、Realsense の alignment ライブラリを使用する。これによって両方のカメラの画角をソフトウェア上で位置合わせが可能になる。

3.2 物体検出のデータセット収集

深層学習による認識ネットワークにはデータセットの数量が多いほど精度とロバスト性が向上する。それは様々な場面での配管の写真を学習することによってどの環境においても対応できる汎用性が高まることを意味している。本研究使用するデータセットの一部を図 3.2 に示す。配管には曲管や T 字管や直管が含まれており、この画像内の中から曲管と T 字管を全て認識できることを目標とする。また、Depth 画像の有効性を示すためにテスト画像では暗闇の中に配管を設置したデータセットを用意した。Depth 画像は光の影響を受けにくいことから、暗闇の中でも配管を認識できるかを検証する。収集したデータはラベリング作業を行う。これは深層学習するにおいての正解データとして、予め画像内のどの部分が曲管又は T 字管であるかをアノテーションする必要がある。本研究では配管画像に対して曲管、T 字管の 2 クラスに分けてラベリング作業を行った。

3.3 6D 姿勢推定のデータセット収集

6D 姿勢推定のデータセットには Colmap を使用して点群データを取得する。Colmap は 2 D 画像から 3 D 点群を再構築するために使用されるソフトウェアである。この 2 D 画像は異なる視点から撮影された同じオブジェクトの画像を複数枚利用することで 3 次元情報を復元することができる。そのため、本研究では曲管と T 字管の周囲をそれぞれ撮影し、Colmap を使用することで点群データを取得した。

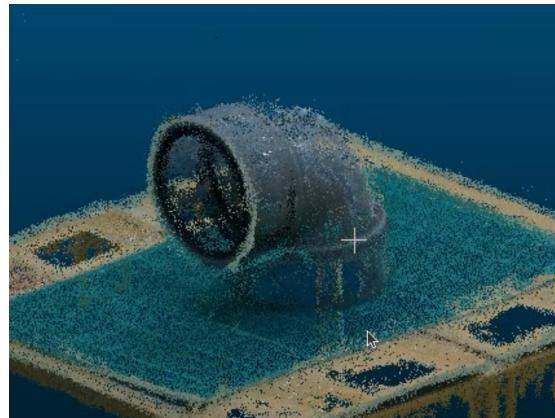


図 3.3: Colmap を用いた曲管の点群データ

第4章 実験

RGB-D カメラから取得したデータセットを RXD ネットワークを使用し曲管又は T 字管を認識できるか検証する。物体認識においては他のネットワークでも実験し、RXD ネットワークの有用性を確かめる。

4.1 評価指標

物体認識の評価指標ではパラメータ数 (Params), Intersection over Union(IoU), mean Average Precision(mAP) を用い認識ネットワークの性能評価を行う。まず、パラメータ数は認識ネットワークの学習可能なパラメータの合計数を示す。これにより、認識ネットワークの複雑度を示すことができる。次に。IoU は正解と予測のバウンディングボックスの共通の重なり部分を 2 つのバウンディングボックスを重ねたときの総面積で除算したものである。IoU は 0 1.0 の値の範囲で示され、値が大きければ大きいほどラベル付されたボックスと予測されたボックスの重なりが正しいことになり、正確に認識していると判断できる。次に、mAP は 1 つ 1 つのクラスに対して平均適合率である AP(Average Precision) を計算する。まず、モデルの予測結果を、出力する信頼度スコア順に並べる。ラベルごとに信頼度スコアがそのラベルの値以上の予測結果について、適合率と再現率を求める。適合率と再現率は図 4.1 のように True Positive(TP) と False Negative(FN) を用いて表される。その適合率と再現率のグラフから適合率の下側の面積を求める。ここで、予測されたラベルが正解なのかの判断は IoU が決められたしきい値以上で、最も信頼度スコアが高い予測ラベルが正解とするように判断される。そして最後に、クラスごとに計算された AP の平均を算出したものが mAP になる。

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}}$$

Intersection
領域の共通部分

Union
領域の和集合

図 4.1: Intersection over Union(IoU)

$$\text{適合率} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{再現率} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

		予測結果	
		曲管	曲管以外
正解	曲管	TP	FN
	曲管以外	FP	TN

図 4.2: 適合率と再現率

4.2 結果と考察

表 4.1: 物体検出ネットワークの実行結果

Network	AP			AP50			Parameters millions
	bent	junction	all	bent	junction	all	
YOLOv3	33.9	68.6	51.3	9.95	20.1	15.0	61.5
YOLOv3-Depth	1.3	0.0	0.7	0.4	0.0	0.2	86.3
RXD	70.9	37.2	54.1	20.8	10.9	15.8	32.4

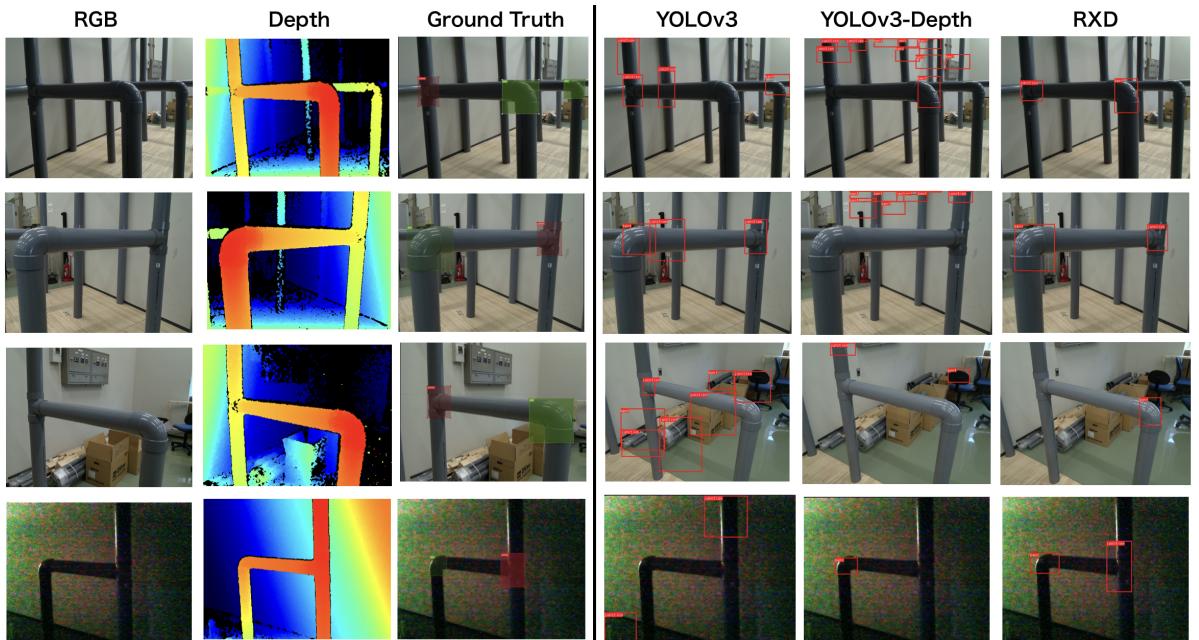


図 4.3: 適合率と再現率

表 4.1 より RXD ネットワークが AP と AP50 の平均値がともに最も高かった。また、パラメータ数に関しては YOLOv3, YOLOv3-Depth よりも低い値となりより優れているネットワークであると言える。RGB 画像だけでなく Depth 画像も学習させると情報量が多くなるため、畳み込む回数も増加しパラメータ数が結果的に多くなる。しかし、RXD ネットワークはパラメータ数を抑えつつ、優れた精度を持っているため実験したネットワークの中で最も良いネットワークであると言える。

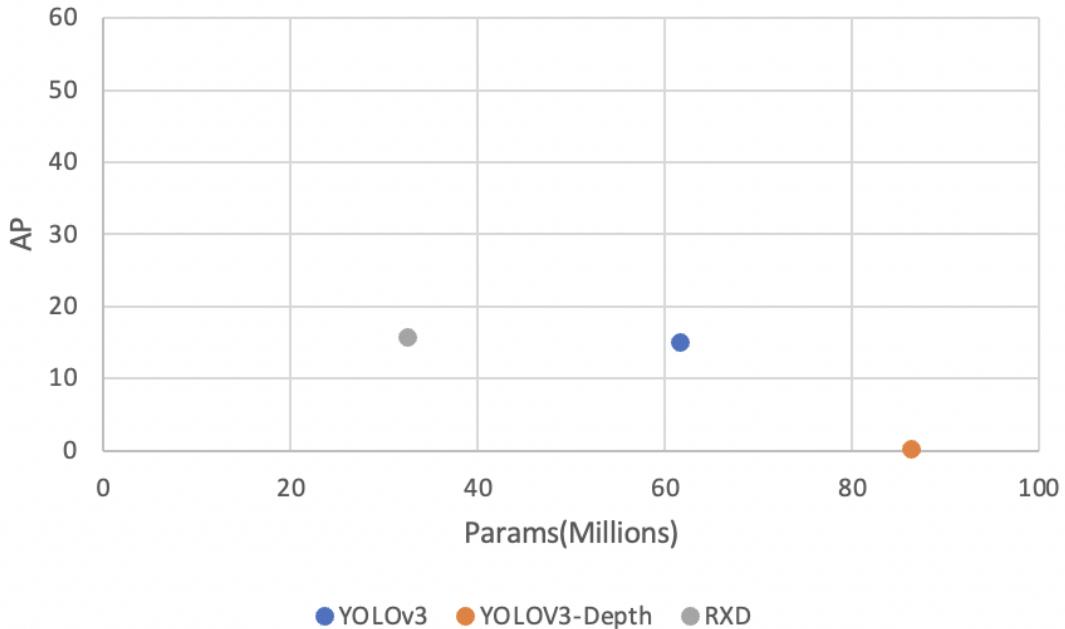


図 4.4: 適合率と再現率

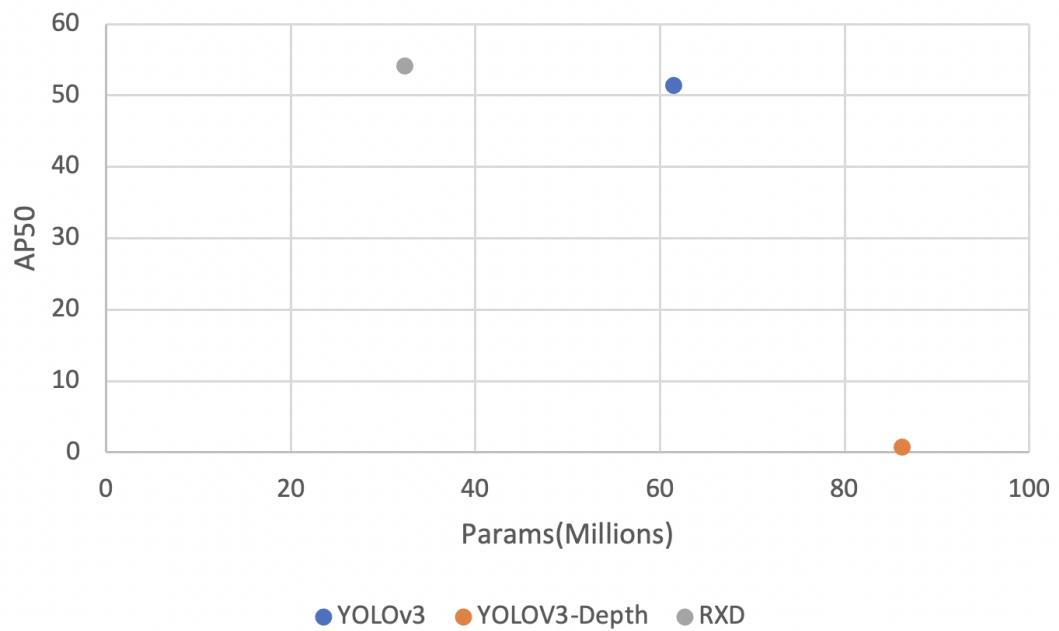


図 4.5: 適合率と再現率

しかし、AP の平均値はともに優れた値であったが、bent と junction 個々の値で見ると YOLOv3 のほうが junction を検出するにおいてはいい結果になっていた。そのため、認識したい物体によってはネットワーク検出器を変更することで、より望ましい結果を得られる可能性もある。評価は AP と AP50 で行ったが、AP のほうが数値が低い結果となった。これは IoU の閾値を上昇させることで認識する条件を厳しくしているため、IoU 閾値を増加させても認識精度が低くならない結果が望ましい。RXD ネットワークは AP の値は AP50 よりも大きく劣っているため、ネットワーク改善を行う必要があると言える。

表 4.2: テスト画像を用いた検出結果

<i>AP</i>	<i>AP50</i>	<i>Parameters</i>
-----------	-------------	-------------------

図4.2の結果にそれぞれのネットワークの出力画像を示した。結果より RXD ネットワークが最も良い検出を示している。しかし、RXD ネットワークの出力されたデータでは T 字管を認識できていない結果も存在している。これは、もとの Depth 画像のデータセットと比較すると遠くの物体になるほどデータが欠落しているため、認識が困難であったと考えられる。そのため、RGB-D カメラの精度が低いと物体検出の精度に影響してくることがわかる。また、暗闇の中でのテスト画像では YOLOv3-Depth と RXD ネットワークの出力結果が配管を認識できていた。これは暗闇の状況下でも影響を受けない Depth 画像が役に立っていると考えられる。

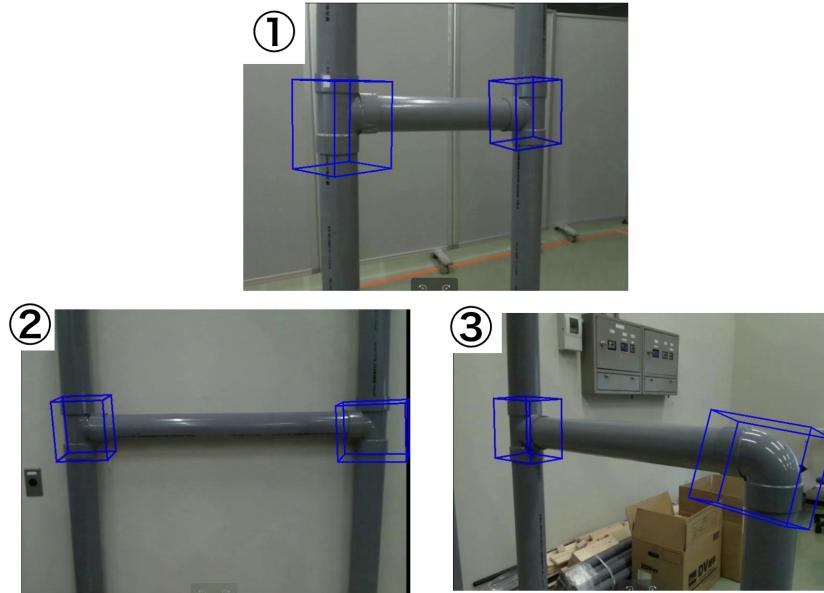


図 4.6: 適合率と再現率

次に、6 D 姿勢推定の結果を図 4.3 に示す。既存の Gen6D のみでは検出器がオブジェクトの複数認識に対応していなかった。RXD ネットワークは画像の中の全てのオブジェクトを認識可能なため、検出された値を Gen6D の Selector に渡すことで複数姿勢推定を可能とする。しかし、結果では曲管の姿勢がボックスとうまく一致しなく望ましくない結果になった。また、図 4.4 のように junction 同士が向かい合っている画像の姿勢推定を行い、それぞれのオブジェクトの Yaw, Pitch, Roll を求めた。表の結より junction 同士が向かい合っていることがわかる。次に、表の結果のそれぞれの姿勢を用いて、Rviz を使用してそれぞれのオブジェクトの座標系を可視化した。完全に向かい合った結果にはならなかったが、T 字管の位置関係と姿勢を表示することができた。

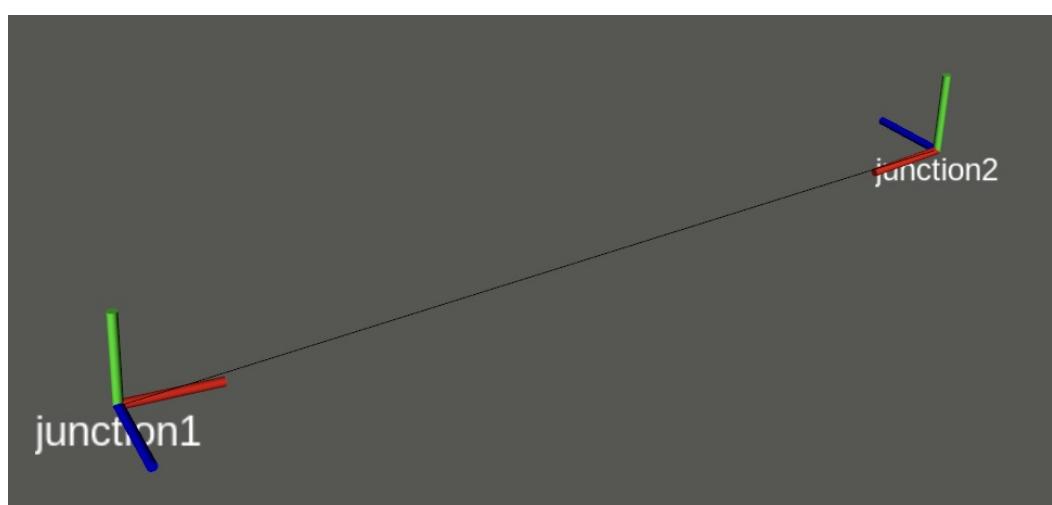


図 4.7: 適合率と再現率

第5章 結言

本稿では、立命館大学 理工学部 ロボティクス学科 川村・金岡研の修士論文、卒業論文執筆の共通化と効率化のために、論文の論理構造、体裁、その他一般の注意点を明文化した。

途中、読み苦しい部分もあったと思われるが、より体系的で読みやすい手引きとなるよう、改訂を重ねて行く予定である。

参考文献

- [1] Author(V. Ferrari, T. Tuytelaars, and L. Van Gool): "Simultaneous object recognition and segmentation from single or multiple model views," *International Journal of Computer Vision* ,vol.67, no.2, pp. 159–188, 2006.
- [2] Author(A. Collet, M. Martinez, and S. S. Srinivasa): "The moped framework: Object recognition and pose estimation for manipulation," *The International Journal of Robotics Research* ,vol.30, no.10, pp. 1284–1306, 2011.
- [3] Author(M. Aubry, D. Maturana, A. A. Efros, B. C. Russell, and J. Sivic): "Seeing 3d chairs: Exemplar part-based 2d-3d alignment using a large dataset of cad models," *in Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)* , pp. 3762–3769, 2014.
- [4] Author(M. A. Fischler and R. C. Bolles): "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM* ,vol.24m no.6, pp. 381–395, 1981.
- [5] Author(S. Tulsiani and J. Malik): "Viewpoints and keypoints," *in Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)* , pp. 1510–1519, 2015.
- [6] Author(M. Schwarz, H. Schulz, and S. Behnke): "Rgb-d object recognition and pose estimation based on pre-trained convolutional neural network features," *in Robotics and Automation (ICRA), 2015 IEEE International Conference on, IEEE* , pp. 1329–1335, 2015.
- [7] Author(Guilhem Cheron, Ivan Laptev, Cordelia Schmid): "P-CNN: Pose-Based CNN Features for Action Recognition," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* , pp. 3218–3226, 2015.
- [8] Author(WANG, Chen): "Densefusion: 6d object pose estimation by iterative dense fusion," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* , pp. 3343–3352, 2019.
- [9] Author(LIU Yuan): "Gen6D: Generalizable model-free 6-DoF object pose estimation from RGB images," *In: Computer Vision–ECCV 2022: 17th European Conference* , pp. 298–315, 2022.

- [10] Author(REDMON, Joseph; FARHADI, Ali): "Yolov3: An incremental improvement," *In: Computer Vision-ECCV 2022: 17th European Conference* , 2018.
- [11] Author(FISHER, Alex): "ColMap: A memory-efficient occupancy grid mapping framework," *Robotics and Autonomous Systems* , 2021.
- [12] Author(GIRSHICK, Ross): "Fast r-cnn," *In: Proceedings of the IEEE international conference on computer vision* , pp. 1440–1448, 2015.
- [13] Author(Xian, Yongqin and Choudhury, Subhabrata and He, Yang and Schiele, Bernt and Akata, Zeynep): "Semantic Projection Network for Zero- and Few-Label Semantic Segmentation," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , pp. 8256–8265, 2019.
- [14] Author(ZHOU, Tao): "RGB-D salient object detection: A survey," *Computational Visual Media* , 7; 37-69, 2021.

付録 A 数式の記述

本付録では、数式の記述例を示す。本手引のクラスファイル `kzthesis.cls` 独自のコマンドの使用法の説明も兼ねているので、参考にすること。

A.1 記述例

まず、受動性の定義を与えておく[?]。システムの入力 u と出力 y が同じ次元であるとする。このとき、ある有限な正の定数 γ_0^2 に対して次式が成り立つならば、システムは受動的であるという。

$$\int_0^t \mathbf{y}^T(\tau) \mathbf{u}(\tau) d\tau \geq -\gamma_0^2, \quad \forall t > 0 \quad (\text{A.1})$$

一方、 n 自由度ロボットの動力学は一般に次のような微分方程式に従う。

$$\mathbf{u} = \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \frac{1}{2}\dot{\mathbf{M}}(\mathbf{q})\dot{\mathbf{q}} + \mathbf{S}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) + \mathbf{d}(\dot{\mathbf{q}}) \quad (\text{A.2})$$

ただし、 $\mathbf{q} \in \mathbb{R}^n$ は関節変位ベクトル、 $\mathbf{u} \in \mathbb{R}^n$ は関節駆動力ベクトルである。右辺第一項は慣性項であり、 $\mathbf{M}(\mathbf{q}) \in \mathbb{R}^{n \times n}$ は実対称正定な慣性行列である。第二、三項は非線形項であり、 $\mathbf{S}(\mathbf{q}, \dot{\mathbf{q}}) \in \mathbb{R}^{n \times n}$ は歪対称行列となる。 $\mathbf{g}(\mathbf{q}) \in \mathbb{R}^n$ はポテンシャル項である。 $\mathbf{d}(\dot{\mathbf{q}}) \in \mathbb{R}^n$ は摩擦等による散逸項であり、その各成分は $\dot{\mathbf{q}}$ の対応する成分と常に同符号である。 $K(\mathbf{q}, \dot{\mathbf{q}})$ 、 $P(\mathbf{q})$ をそれぞれ運動エネルギー、ポテンシャルエネルギーとすると、以下のような関係が成り立つ。

$$K(\mathbf{q}, \dot{\mathbf{q}}) = (1/2) \dot{\mathbf{q}}^T \mathbf{M}(\mathbf{q}) \dot{\mathbf{q}} \quad (\text{A.3})$$

$$\mathbf{g}(\mathbf{q}) = (\partial P(\mathbf{q}) / \partial \mathbf{q}^T)^T \quad (\text{A.4})$$

このとき、ロボットの全内部エネルギーは $E(\mathbf{q}, \dot{\mathbf{q}}) = K(\mathbf{q}, \dot{\mathbf{q}}) + P(\mathbf{q})$ となる。

ロボットシステムは、速度出力 $y = \dot{\mathbf{q}}$ に関して受動的であることが知られている。すなわち、式 (A.2) から、

$$\begin{aligned} \int_0^t \dot{\mathbf{q}}^T \mathbf{u} d\tau &= \int_0^t \dot{\mathbf{q}}^T \left\{ \mathbf{M}\ddot{\mathbf{q}} + \frac{1}{2}\dot{\mathbf{M}}\dot{\mathbf{q}} + \mathbf{S}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) + \mathbf{d}(\dot{\mathbf{q}}) \right\} d\tau \\ &= \int_0^t \frac{d}{d\tau} \{K(\mathbf{q}, \dot{\mathbf{q}}) + P(\mathbf{q})\} d\tau + \int_0^t \dot{\mathbf{q}}^T \mathbf{d}(\dot{\mathbf{q}}) d\tau \\ &= E(\mathbf{q}(t), \dot{\mathbf{q}}(t)) - E(\mathbf{q}(0), \dot{\mathbf{q}}(0)) + \int_0^t \dot{\mathbf{q}}^T \mathbf{d}(\dot{\mathbf{q}}) d\tau \\ &\geq -E(\mathbf{q}(0), \dot{\mathbf{q}}(0)) = -\gamma_0^2 \end{aligned} \quad (\text{A.5})$$

となり、式 (A.1) が成立する。ただし、 S の歪対称性と $\dot{\mathbf{q}}^T \mathbf{d}(\dot{\mathbf{q}}) \geq 0$ を用いた。上式から、 γ_0^2 はロボットの初期内部エネルギーと解釈できる。