

P D III

## 蛍光ARマーカを用いた6次元姿勢推定

指導教員 出村 公成 教授



金沢工業大学  
工学部ロボティクス学科

川隅 蓮

令和3年度  
2022年2月3日

# 目次

<b>第1章 序 論</b>	<b>1</b>
1.1 はじめに . . . . .	1
1.2 本論文の構成 . . . . .	2
<b>第2章 従来研究</b>	<b>3</b>
2.1 アノテーションにおける従来研究 . . . . .	3
2.2 蛍光塗料を活用した従来研究 . . . . .	3
<b>第3章 提案手法</b>	<b>5</b>
3.1 概 要 . . . . .	5
3.2 萤光ARマーカ作成 . . . . .	5
3.3 ハードウェア . . . . .	5
3.4 データ拡張 . . . . .	6
3.5 萤光ARマーカ認識 . . . . .	7
3.6 姿勢推定 . . . . .	8
3.7 位置推定 . . . . .	9
<b>第4章 実 験</b>	<b>10</b>
4.1 環 境 . . . . .	10
4.2 姿勢推定に対応する実験 . . . . .	10
4.2.1 実験方法 . . . . .	10
4.2.2 実験結果及び考察 . . . . .	10
4.3 位置推定に対応する実験 . . . . .	11
4.3.1 実験手法 . . . . .	11
4.3.2 実験結果及び考察 . . . . .	11
<b>第5章 結 言</b>	<b>14</b>
5.1 終わりに . . . . .	14

5.2 今後の課題 . . . . .	14
謝 辞	16
参考文献	17
本研究に関する学術発表論文	19

# 図目次

3.1	Fluorescence AR Marker . . . . .	6
3.2	Environment . . . . .	6
3.3	Normal light and UV light . . . . .	7
3.4	Dynamixel MX-106T . . . . .	7
3.5	Hardware overview . . . . .	8
3.6	Threshold image . . . . .	8
3.7	Segmentation image . . . . .	9
4.1	Estimation error with a same background than when learning . . . . .	12
4.2	Estimation error with a different background than when learning . . . . .	12
4.3	One of the data sets . . . . .	13
4.4	Segmentation error . . . . .	13

# 表 目 次

4.1	About marker . . . . .	10
4.2	Average and standard deviation with the same background as when learning . . .	11
4.3	Average and standard deviation with a different background than when learning .	11
4.4	Average and standard deviation . . . . .	13

# 第1章

## 序論

### 1.1 はじめに

近年、深層学習を用いた認識が注目されているが、高い精度の学習を行うには大量のデータセットが必要になり、そのアノテーション作業のコストが深層学習の課題の1つになっている。アノテーションの自動化として、シミュレーションを使った手法や、特殊なデバイスや環境を使った手法がある。シミュレーションを使った手法では、現実とシミュレーション環境のクオリティの違いが問題となっている。特殊なデバイスや環境を使った手法では、紫外線を照射されると発光する蛍光塗料を使い、対象物体のアノテーションが行われている。岡野[1]らの手法では、蛍光塗料を塗ったARマーカやグリッド状のテクスチャを不定形物体に貼ることで、3次元的な位置姿勢についてアノテーションを行っているが、紫外線強度とマーカ認識精度の関係などの基礎的な研究が主で、その後の位置姿勢推定などの応用はされていない。

本研究では蛍光ARマーカ用いてデータセット作成を行い、3Dモデルに依存しない6次元位置姿勢推定手法を開発した。蛍光ARマーカは、紫外線下で発光する塗料を透明フィルムに塗り、ARマーカのテクスチャに合わせて切り抜いたものである。可視光下では透明であり対象物体のテクスチャを保持したまま、6次元位置を取得することができる。紫外線と可視光を切り替えて撮影することで、RGB画像・UV画像・6次元位置姿勢を1つのデータセットとして取得できる。本研究の目的は、3Dモデルを持たない生活品をロボットが把持するために、低成本でデータセット作成ができ、マニピュレーションなどのロボットタスクに必要6次元姿勢を推定することである。

## 1.2 本論文の構成

本論文の構成について述べる。2章では、これまでの自動アノテーションと蛍光塗料を活用した研究について述べる。3章では、本研究で提案する蛍光ARマーカによる自動アノテーション手法と、6次元姿勢推定について述べる。4章では位置と姿勢の推論精度を検証する。5章では本論文の結論をまとめ、今後の展望について述べる。

## 第2章

### 従来研究

#### 2.1 アノテーションにおける従来研究

アノテーションは非常に時間がかかる作業であり、効率化のために人の介入をなくした自動アノテーションが研究されている。自動アノテーションでは、シミュレーションを使った手法や特殊なデバイスや環境を使った手法がある。

シミュレーションを使った手法 [2,3] では、対象物体の 3D モデルをインポートした環境で、光の当たり方や位置姿勢を自動で変化させデータセットを作成する。短時間で大量のデータセットを作成できるが、日用品などの 3D モデルが手に入りづらい物体への応用や、現実とシミュレーション環境のクオリティの違いが問題となっている。

特殊なデバイスや環境を使った手法 [1,4–8] では、人追従ロボットに取り付けられた引き紐を人が引くことで、ロボットが進むべき方向をアノテーションしている [4]。認識では、AR マーカを対象物体に貼りつけ自動アノテーションをする研究がある。AR マーカは 6 次元姿勢情報とマーカ固有 ID が取得できるが、マーカが貼られた箇所が隠れてしまうオクルージョン問題がある。オクルージョンの解決として、AR マーカを対象物体ではなく台座に貼った研究もある [5,6]。

#### 2.2 蛍光塗料を活用した従来研究

K.Takahashi らの手法 [7] は、蛍光塗料を塗った対象物体に可視光と紫外線を切り替えて撮影することで、蛍光塗料の領域を抽出したマスク画像を作成した。可視光と紫外線で撮影した画像の差分からマスク画像を得ることで、光がコントロールされた特殊な環境を必要としない。

岡野[1]らの手法では、蛍光塗料を塗ったARマーカやグリッド状のテクスチャを対象物体に貼り付け、K.Takahashiらの手法では得られなかった3次元的な位置と姿勢をアノテーションしている。また、蛍光塗料は可視光では透明でありオクルージョン問題を解決している。紫外線強度と蛍光ARマーカの位置姿勢精度の関係や、マーカテクスチャとグリッドテクスチャの精度比較など基礎的な研究が主で、6次元姿勢推定の学習などの応用は研究されていない。本研究では、蛍光ARマーカで作成したデータセットで6次元姿勢推定を開発し、マニピュレーションなどのロボットタスクでの有用性を評価する。

# 第3章

## 提案手法

### 3.1 概要

本節では、本研究で提案するシステム全体についての概要を述べる。認識させたい対象物体に蛍光 AR マーカを貼り、通常光と紫外線で撮影することで、RGB 画像,UV 画像,6 次元情報を得られる。これらの情報から位置と姿勢を学習させて、RGB 画像の入力から位置姿勢を推論させる。データセットの作成は光がコントロールされた環境で行う(図 3.2)。

### 3.2 蛍光 AR マーク作成

AR マーカの出力には ROS の Ar-track-alvar を用い、マーカのサイズは  $50[\text{mm}] \times 50[\text{mm}]$  にする。AR マーカの作成は、蛍光塗料を塗った透明フィルムをレーザー加工機で切り抜き、一部をカッターナイフで修正した(図 3.1)。

### 3.3 ハードウェア

データセット作成は人の介入を減らすために、可視光と紫外線の切り替えと、対象物体の回転を自動化した(図 3.3)。撮影するカメラには RealSense D435 を使用し、対象物体までの距離は 300mm~600mm とする。ハードウェアの構成を図 3.5 に示す。可視光と紫外線の切り替えには、AC100V の ON/OFF 可能なソリッドステートリレー(SSR)を、Arduino Uno で制御している。対象物体は Dynamixel MX-106T を使った回転台に載せて  $10[^\circ]$  ごとに回転させた(図 3.4)。PC(ROS melodic)から Arduino Uno へ撮影開始信号が送られると、可視光が照射され 1 枚の画像を撮影する。可視光から紫外線へ切り替え、1 枚の画像を撮影する。対象物体を載せた回転台が  $10[^\circ]$  回転する。この一連の流れには約 5 秒かかる。



Fig. 3.1: Fluorescence AR Marker



Fig. 3.2: Environment

### 3.4 データ拡張

画像には以下に列挙するデータ拡張手法を適用した。

- Contrast transformation
  - Gamma correction
  - Blur
  - Gaussian noise
  - Histogram equalization



Fig. 3.3: Normal light and UV light

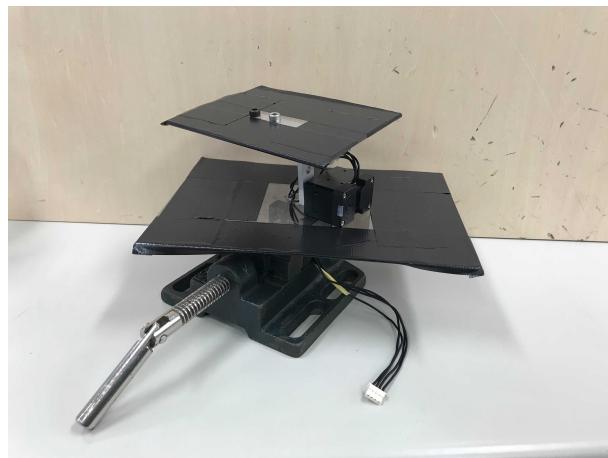


Fig. 3.4: Dynamixel MX-106T

### 3.5 荧光ARマーカ認識

AR マーカの認識アルゴリズムでは、ペイロードが黒と白で表現されていることを前提としているため、紫外線下で撮影した画像を認識させることはできない。そのため紫外線下で撮影した RGB 画像を二値化した。二値化には適応的二値化を用いた(図 3.6)。適応的二値化は画像全体で 1 つのしきい値を推定する大津二値化と異なり、画像を局所領域に分割し、その中でしきい値を推定する。大津二値化に比べてカメラと対象物体の距離が変化した場合や、外光の変化に対して影響を受けづらい。RealSenseD435 の画像サイズ 640 x 480 に対して、適応的二値化の局所領域を 101 x 101 とした。

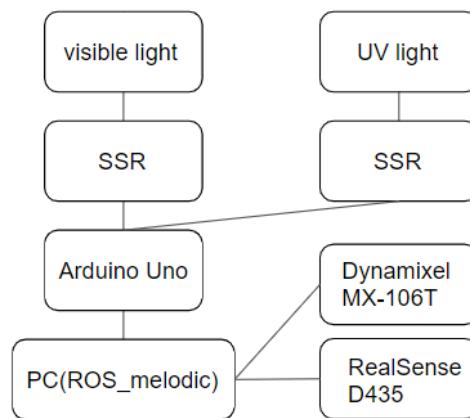


Fig. 3.5: Hardware overview

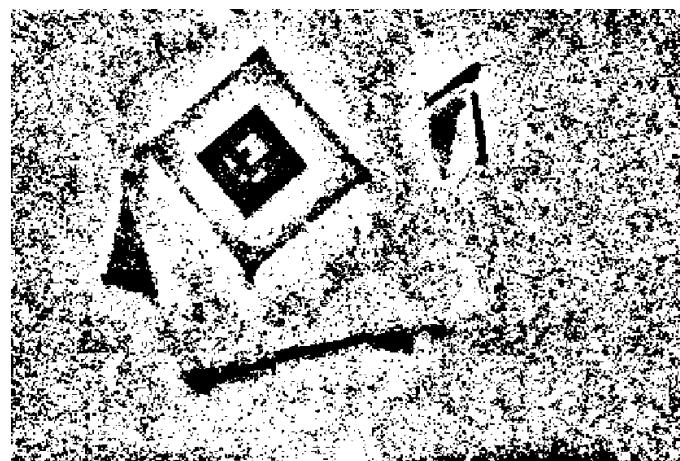


Fig. 3.6: Threshold image

## 3.6 姿勢推定

姿勢推定では、対象物体の映った RGB 画像を入力して Roll,Pitch,Yaw を出力する。ネットワークモデルには ResNet [9] を用いる。損失関数には平均二乗誤差を用いた。学習率は  $n=0$  を 20epoch ごとにインクリメントさせ、以下の通りに示される。

$$lr = 0.98^n \quad (3.1)$$

学習データセットは、1036 セットをデータ拡張で 8288 セットに拡張した。ミニバッチサイズは 4、epoch 数は 200、GTX1080Ti で学習には約 20 時間かかった。

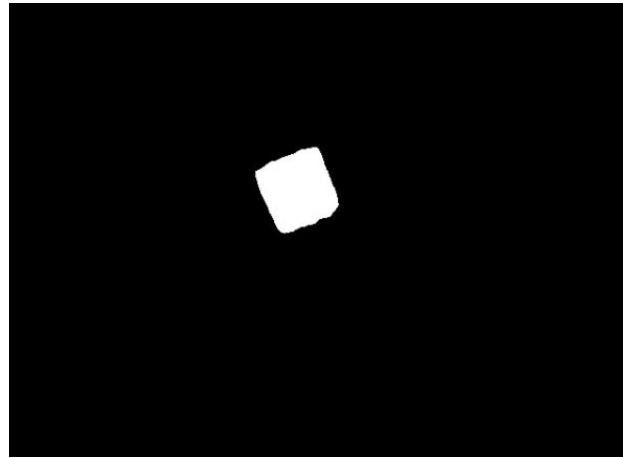


Fig. 3.7: Segmentation image

### 3.7 位置推定

位置推定では、対象物体の映った RGB 画像を入力して、セマンティックセグメンテーションを行う。フレームワークには segmentation-models-pytorch を使用し、UV 画像の蛍光塗料部分を色抽出したマスク画像を教師データとして、Unet [10] で学習する。学習データセットは 783 セット。ミニバッチサイズは 4、epoch 数は 40、GTX1080Ti で学習には約 1 時間かかった。セマンティックセグメンテーションされた画像(図 3.7)は、セグメンテーション部分の重心のピクセルから、深度カメラ Realsense D435 から 3 次元座標を求める。計算式は以下の通りに示される。

$$\text{HorizontalAngle} = (px - \frac{W}{2}) \cdot \frac{HFOV}{W} \quad (3.2)$$

$$\text{VerticalAngle} = (py - \frac{H}{2}) \cdot \frac{VFOV}{H} \quad (3.3)$$

$$x = depth \cdot \tan(\text{HorizontalAngle}) \quad (3.4)$$

$$y = depth \cdot \tan(\text{VerticalAngle}) \quad (3.5)$$

$px$  と  $py$  はセグメンテーション部分の重心のピクセル、 $W$  と  $H$  は Realsense D435 の画像サイズ、 $HFOV$  と  $VFOV$  は Realsense D435 の画角である。

# 第4章

## 実験

### 4.1 環境

本手法では外光が入らず、通常の可視光と紫外線を照射できるコントロールされた環境を利用する(図3.2,表4.1)。紫外線の照射には東芝4形ブラックライト直管蛍光灯を用いた。

### 4.2 姿勢推定に対応する実験

#### 4.2.1 実験方法

実験では、RGB画像からRoll,Pitch,Yawを推論させて、教師データとの平均誤差[°]と標準偏差[°]を求める。推論には、学習時と同じ背景で撮影した15枚の画像と、学習時とは異なる背景で撮影した9枚の画像を用いる。教師データは蛍光ARマーカから取得したRoll,Pitch,Yawとする。

#### 4.2.2 実験結果及び考察

学習時と同じ背景で撮影したRGB画像の推論結果を図4.1と表4.2に示す。Yawは平均誤差[°]と標準偏差[°]が、3[°]以下となっており、従来のARマーカの代替えとして利

Table 4.1: About marker

ARMarker	ar-track-avlvar
ARMarkerSize	50 [mm]
カメラとマーカの距離	300～600 [mm]

Table 4.2: Average and standard deviation with the same background as when learning

	average[°]	standard deviation[°]
Roll	10.4	4.43
Pitch	9.88	6.54
Yaw	2.85	2.19

Table 4.3: Average and standard deviation with a different background than when learning

	average[°]	standard deviation[°]
Roll	8.39	7.93
Pitch	11.4	6.46
Yaw	8.57	13.1

用可能と考えれる。Roll と Pitch は平均誤差と標準偏差が大きい。原因として、Roll と Pitch はデータセット作成時に、マーカとして認識できる範囲が狭いため、Yaw に比べてバリエーションが少ないデータセットが作られたと考える。図 4.3 に Roll が 25.4[°] とデータセット内では最大値の教師画像を示す。

学習時とは異なる背景で撮影した RGB 画像の推論結果を図 4.2 と表 4.3 に示す。学習時と同じ背景で推論させた結果と比べて、Roll, Pitch, Yaw で平均誤差と標準偏差が大きくなつた。現在の学習方法では、背景も考慮して学習されているからだと考えられる。今後は、蛍光 AR マーカが貼られてた対象物体のみ着目する手法を検討したい。

## 4.3 位置推定に対応する実験

### 4.3.1 実験手法

実験では、紫外線を照射した環境に蛍光 AR マーカが貼られた対象物体を置き、位置 X, Y, Z を計測する。次に、可視光に切り替えて対象物体を計測する。この作業を 10 回繰り返し、蛍光 AR マーカと推論値の差の平均誤差 [mm] と標準偏差 [mm] を求める。

### 4.3.2 実験結果及び考察

推論結果と教師データの平均誤差 [mm] と標準偏差 [mm] を図 4.4 と表 4.4 に示す。x と y は平均誤差と標準偏差が約 10[mm] 程度と、従来の AR マーカの代替えとなれる精度と

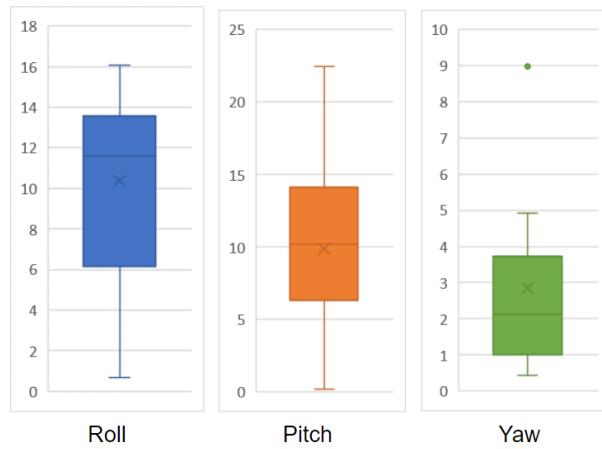


Fig. 4.1: Estimation error with a same background than when learning

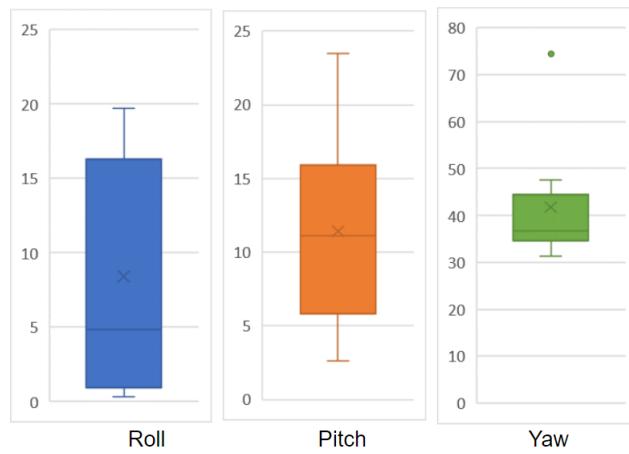


Fig. 4.2: Estimation error with a different background than when learning

なったと思われる。Zは平均誤差が18[mm]とXとYに比べると誤差が大きくなかった。ZはRealsense D435から得られたデプス情報としているため、今回の手法とは関係がないため、より高精度な深度カメラを使うことで誤差は解決できると考えられる。

現在は、セマンティクセグメンテーションによって対象物体を認識させているため、対象物体が複数存在した場合は区別ができない。今後はインスタンスセグメンテーションを実装して、複数物体の認識にも対応させたい。

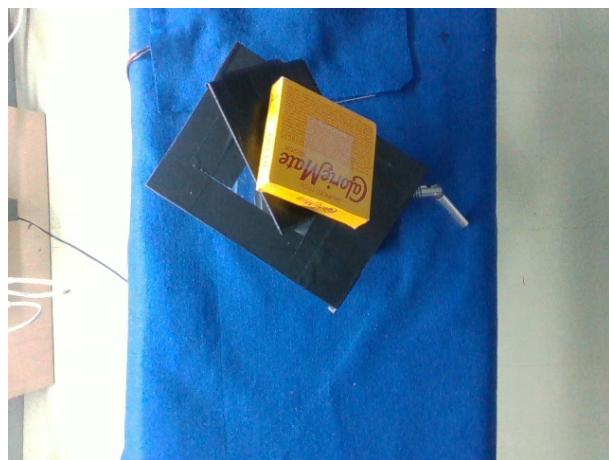


Fig. 4.3: One of the data sets

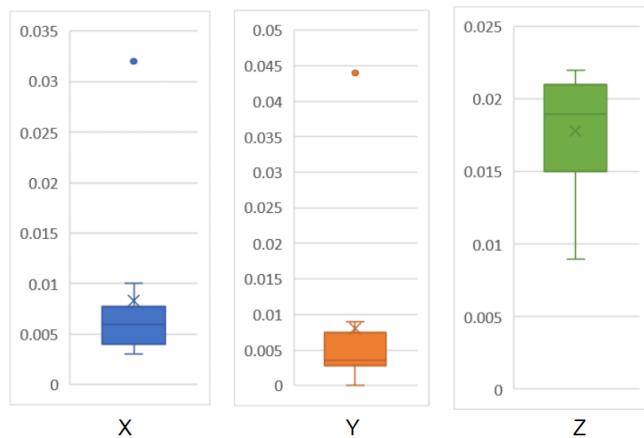


Fig. 4.4: Segmentation error

Table 4.4: Average and standard deviation

	average[mm]	standard deviation[mm]
X	8.3	8.5
Y	8	13
Z	18	4

# 第5章

## 結 言

### 5.1 終わりに

本研究では、ARマーカのテクスチャに蛍光塗料を塗った蛍光ARマーカでアノテーションを可能にし、6次元姿勢推定を行った。データセットの作成の作業を自動化したことで、深層学習の問題の1つであるアノテーション作業のコストを減らすことができた。

姿勢推定では、蛍光ARマーカから得られた姿勢を教師データとして、ResNetによってRoll,Pitch,Yawを推論させた。学習時の背景と同じRGB画像では、Yaw軸で平均誤差[°]と標準偏差[°]が3[°]以下となったが、RollとPitchでは十分な精度が得られなかった。また学習時と背景が異なると、平均誤差がすべての角度で大きくなってしまった。

位置推定では、蛍光ARマーカから得られた位置を教師データとして、セマンティックセグメンテーションを学習させた。位置X,Y,Zで平均誤差と標準偏差が20[mm]以下となり、従来のARマーカの代替えとなれる精度になった。

### 5.2 今後の課題

姿勢推定のRoll,PitchはYawに比べて、ARマーカとして認識できる範囲が狭いため作成されるデータセットで、バリエーションが少なく精度の高い推論を行うことができなかつた。今後は、蛍光ARマーカの可視光下で透明である利点を活かして、対象物体の文字や絵などのテクスチャを利用して、対象物体の向きを効率的に学習できるネットワークを開発したい。また背景の変化で精度が低下してしまうことから、蛍光ARマーカが貼られた付近のみに着目して学習する必要がある。

位置推定では、対象物体が複数存在した場合に個別に位置が推論させるために、インスタンスセグメンテーションを実装したい。

最後に、陳列作業などのマニピュレーションに応用することで、本手法と既存の手法の比較を行いたい。

## 謝 辞

本研究を行うにあたり全体を通してご指導、ご教授、議論などのご助力をいただきました本学ロボティクス学科の出村公成教授に深く感謝いたします。

令和4年2月

## 参考文献

- [1] 岡野昭伍, 牧野達仁, 出村公成. 蛍光テクスチャによる深層学習のための三次元情報を含む自動アノテーション手法の提案. 日本ロボット学会論文誌, Vol. 40, No. 1, 2022.
- [2] 内海佑斗, 和田健太郎, 岡田慧, 稲葉雅幸. 物体メッシュモデルを用いた学習データ自動生成に基づく透明物体の深度画像予測と家事支援ロボットへの応用. 人工知能学会全国大会論文集, Vol. JSAI2018, No. 0, pp. 3L202–3L202, 2018.
- [3] Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, Yunfei Bai, Matthew Kelcey, Mrinal Kalakrishnan, Laura Downs, Julian Ibarz, Peter Pastor, Kurt Konolige, et al. Using simulation and domain adaptation to improve efficiency of deep robotic grasping. pp. 4243–4250, 2018.
- [4] 岡田真也, 上田隆一, 林原靖男. 引き紐を利用した視覚と行動の end-to-end 学習による移動ロボットの人追従行動の生成. 人工知能学会全国大会論文集, pp. 1082–1086, 2019.
- [5] Takuya Kiyokawa, Keita Tomochika, Jun Takamatsu, and Tsukasa Ogasawara. Fully automated annotation with noise-masked visual markers for deep-learning-based object detection. *IEEE Robotics and Automation Letters*, Vol. 4, No. 2, pp. 1972–1977, Feb 2019.
- [6] Takuya Kiyokawa, Keita Tomochika, Jun Takamatsu, and Tsukasa Ogasawara. Efficient collection and automatic annotation of real-world object images by taking advantage of post-diminished multiple visual markers. *Advanced Robotics*, Vol. 33, No. 24, pp. 1264–1280, 2019.
- [7] Takahashi Kuniyuki and Yonekura Kenta. Invisible marker: Automatic annotation of segmentation masks for object manipulation. *International Conference on Intelligent Robots and Systems*, 2020.

- [8] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *International journal of computer vision*, Vol. 92, No. 1, pp. 1–31, 2011.
- [9] He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian. Deepresidual learning for image recognition. *CVPR*, 2016.
- [10] Ronneberger Olaf, Fischer Philipp, and Brox Thomas. U-net: Convolutional networks for biomedical image segmentation. *CVPR*, 2015.

## 本研究に関する学術発表論文