画像生成ネットワークを用いた反復処理に基づく姿勢推定に関する研究

TR18005 大西 剛史　　指導教授：藤吉 弘亘

1. はじめに

ロボットによる物体把持は，把持対象である物体の 6D 姿勢を求める必要がある．従来の姿勢推定手法の入力には，点群を用いる [1]．しかしながら，物体の材質の影響により点群の欠損が生じ，姿勢推定が困難になる場合がある．また，画像を用いた手法 [2] は，テクスチャの変化，光源等によるアピアランスの変化や 1 回の処理しか行わないことによる大きな推定誤差が生じることがあり，把持に失敗することがある．そのため，点群の欠損やアピアランスの変化に依存せず，反復処理を用いた安定した姿勢推定が必要となる．そこで本研究では，セマンティックセグメンテーションと画像生成ネットワークを用いて，誤差逆伝播を繰り返し行う姿勢推定手法を提案する．

2. 従来の姿勢推定

物体の姿勢推定では，世界座標系の x, y, z 軸に関する回転と並進量を求める．物体の姿勢推定手法には，モデルベース手法 [1] とアピアランスベース手法 [2] がある．モデルベース手法は，3D モデルとセンサから取得した点群のマッチングにより姿勢推定を行う．代表的なモデルベース手法として ICP[1] が提案されており，点群を用いた反復処理により高い推定精度を達成している．アピアランスベース手法は，あらかじめ多視点から撮影した画像を学習し，対象画像のアピアランスを照合することで姿勢の推定を行う．また，Convolutional Neural Network(CNN) を用いた手法 [2] が提案されており，画像座標系における物体の中心座標やカメラの姿勢を推定することで 1 回の処理による高い推定精度を達成している．

3. 提案手法

提案手法は，姿勢パラメータを入力してセグメンテーション画像を生成するネットワークに生成結果と目標画像の誤差を逆伝播することで，姿勢パラメータの更新量を求める．ここでの姿勢パラメータは回転と並進移動量である．そして，更新量を加えた姿勢パラメータを入力することで，再度更新量を得る．これを繰り返すことで正しい姿勢を推定する．図 1 に示すように提案手法は，実画像に対するセグメンテーション，画像生成ネットワークによるセグメンテーション画像生成，誤差逆伝播による姿勢更新から構成される．

実画像に対するセグメンテーションは，テクスチャの変化や光源等の照明変化による姿勢推定への影響を抑制した目標画像の生成を目的とする．画像生成ネットワークは，姿勢パラメータを入力に与え，入力した姿勢でレンダリングする物体のセグメンテーション結果を出力する．ここで，レンダリングに画像生成ネットワークを用いるのは，中間層が姿勢パラメータから画像生成をするための重みを多く持ち，姿勢更新に寄与すると考えるためである．

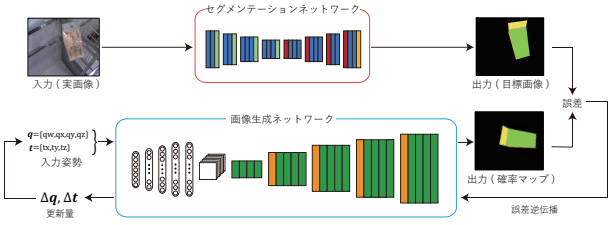


図 1：提案手法の流れ

3.1 物体のセグメンテーション

姿勢パラメータの推定に用いる誤差を計算する際，物体のテクスチャの変化や光源の位置，強度等の撮影環境の変化，背景の影響を低減するために，実画像のセグメンテーションを行い，目標画像を生成する．本研究で対象とするサンドイッチは，図 2 に示すように右側面，背面，底面，耳部，左側面，正面，背景の 7 クラスとなる．面だけでは正面のみしか見えない場合，姿勢推定時に上下の判定が困難になる．そこでパーツとして耳部のクラスを定義する．セグメンテーションネットワークには，Encoder-Decoder 構造で画素単位のクラス識別を行う SegNet[3] を用いる．

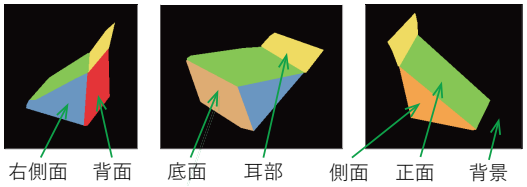


図 2：サンドイッチにおける面ラベルの定義

3.2 姿勢の更新に用いる画像生成ネットワーク

画像生成ネットワークは入力に姿勢パラメータを与え，複数解像度における画素ごとの各クラス確率を出力する．姿勢パラメータは物体の回転を表すクォータニオン $\boldsymbol{q} = (q_w, q_x, q_y, q_z)$ と並進ベクトル $\boldsymbol{t} = (t_x, t_y, t_z)$ の 7 次元ベクトルとする．ネットワーク構造は図 3 とする．この入力に対する特徴を全結合層で得たのち，Upsampling Block により空間的な情報を捉えつつ特徴マップを拡大する．ここで，Upsampling Block は畳み込み層と Unpooling によって構成されている．それぞれ 20×20，40×40，80×80，160×160，320×320 の解像度の特徴マップを出力し，各解像度におけるソフトマックスクロスエントロピー誤差の平均を学習誤差とする．低解像度の誤差を学習誤差に含めることで高解像度の生成時に輪郭の欠損を抑制する．

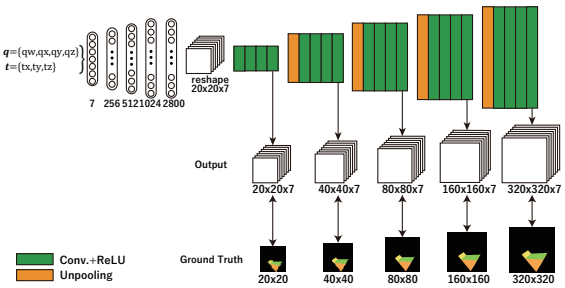
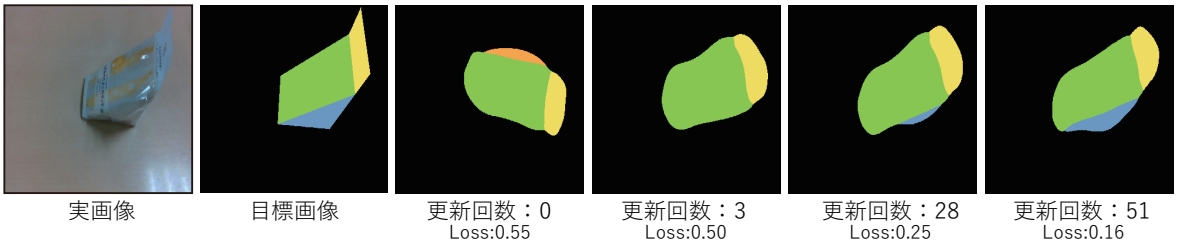


図 3：画像生成ネットワークの構造

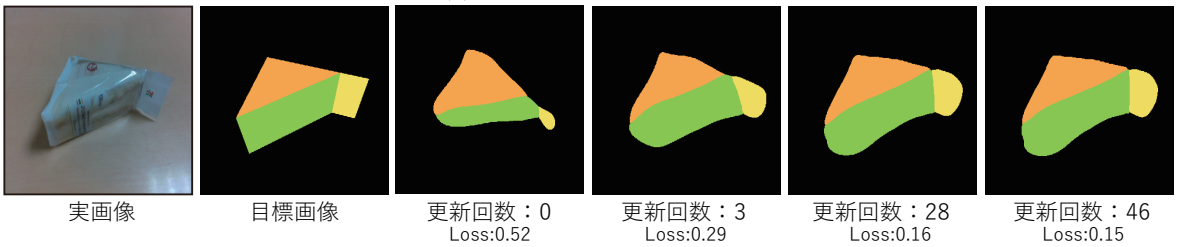
3.3 誤差逆伝播による姿勢推定

提案手法は，画像生成ネットワークの出力とセグメンテーション結果の誤差を求め，画像生成ネットワークに逆伝播することで姿勢パラメータを更新する．姿勢パラメータの更新に用いる誤差は，式 (1)(2) により求める．式 (2) は，画素ごとに教師のクラスと推定したクラスを確認し，教師と推定したクラスが共に背景の場合は 0，それ以外は正解クラスの確率が 1 に近づくように誤差を計算している．ここで，背景クラスのクラスラベルは 0 とする．式 (1) は式 (2) の誤差が 0 以外の画素数で割って平均を求めている．ここで，$\boldsymbol{P}$ は画像生成ネットワークの出力，$\boldsymbol{C}$ は目標のセグメンテーションラベル，k はクラス数，$\boldsymbol{p} = \{p_0, p_1, \cdots, p_k\}$ は 1 画素に対する各クラス確率，c は正解クラスを表し，c が 0 のとき，p_c は p_0 となる．また，N は全画素数，n は p_c, c が共に背景ラベルの場合の画素数を表す．

$$L(\boldsymbol{P}, \boldsymbol{C}) = \frac{\sum_{\{\boldsymbol{p}\in\boldsymbol{P},\ c\in\boldsymbol{C}\}}^{N} l(\boldsymbol{p}, c)}{N - n} \tag{1}$$



(a) 縦置き姿勢シーンの推定例

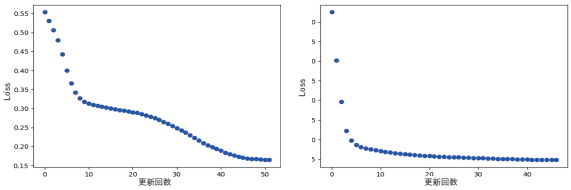


(b) 横置き姿勢シーンの推定例

図 4：提案手法の姿勢推定例

表 1：推定誤差の比較

データ	手法	指標	$r_x[deg]$	$r_y[deg]$	$r_z[deg]$	$t_x[mm]$	$t_y[mm]$	$t_z[mm]$
点群	ICP	平均誤差	41.38	26.54	94.23	**7.18**	**5.29**	**3.38**
		分散	± 1446.12	± 332.58	± 2690.38	**± 023**	**± 0.41**	**± 0.31**
画像	6D-PoseNet	平均誤差	**26.66**	27.19	37.92	34.99	29.67	72.98
		分散	± 2337.45	± 489.08	± 4327.06	± 1.82	± 1.32	± 3.86
	提案手法	平均誤差	30.41	**23.11**	**17.62**	10.48	9.19	20.68
		分散	**± 682.04**	**± 200.53**	**± 1101.65**	± 0.43	± 0.46	± 1.12



(a) 縦置き姿勢シーンの誤差の推移 (b) 横置き姿勢シーンの誤差の推移

図 5：誤差の推移

$$l(\boldsymbol{p}, c) = \begin{cases} 0 & \text{if} \arg\max(\boldsymbol{p}) = 0 \cap c = 0 \\ 1 - p_c & \text{otherwise} \end{cases} \quad (2)$$

誤差逆伝播によって得られる姿勢パラメータの勾配 $\Delta \boldsymbol{q}, \Delta \boldsymbol{t}$ を求め，この勾配を用いて，姿勢パラメータを更新する．勾配計算と姿勢パラメータ更新は，式 (3) に示す．ここで，α は更新率を表す．

$$\begin{cases} \boldsymbol{q} \leftarrow \boldsymbol{q} - \alpha \Delta \boldsymbol{q}, \Delta \boldsymbol{q} = \frac{\partial L(\boldsymbol{P},\ \boldsymbol{C})}{\partial \boldsymbol{q}} \\ \boldsymbol{t} \leftarrow \boldsymbol{t} - \alpha \Delta \boldsymbol{t}, \Delta \boldsymbol{t} = \frac{\partial L(\boldsymbol{P},\ \boldsymbol{C})}{\partial \boldsymbol{t}} \end{cases} \quad (3)$$

誤差逆伝播によって姿勢パラメータを繰り返し更新し，生成画像とセグメンテーション画像が一致するまで行う．反復処理の終了条件は，誤差の変動が閾値よりも小さく収束したときとする．

4. 評価実験

提案手法が有効であることを示すために，点群を用いた ICP とセグメンテーション画像から直接姿勢を推定する 6D-PoseNet を比較対象とする．また，初期姿勢として入力する姿勢パラメータは，回転は 20 通り，並進は目標画像の物体の重心座標を逆透視投影変換した値を用いて，式 (1) により計算した誤差が最も小さい姿勢とする．

4.1 評価方法

評価では，x, y, z 軸それぞれに関する回転誤差と並進誤差を比較する．また，姿勢推定の安定性を確認するため，それぞれの誤差の分散を示す．

4.2 実験結果

評価結果を表 1 に示す．また，姿勢推定過程を図 4 に示す． 図 4 より更新回数が増加するにつれ，目標の画像に近似することを確認できる．これにより，画像生成ネットワークを用いた反復処理は，姿勢推定に効果があると言える．ここで，図 4(a)(b) の誤差の推移は，それぞれ図 5(a)(b) のようになっており，更新回数が増加するにつれ，誤差が低下していることが確認できる．また，図 4(a)(b) は異なるテクスチャの商品である．提案手法はセグメンテーションを目標にしているため，テクスチャが異なる商品に対しても有効であることが確認できる．表 1 の ICP と提案手法の回転誤差の平均を比較すると，提案手法は点群を用いた ICP と比べ，回転誤差が小さいことがわかる．また，回転誤差の分散を見ると提案手法の方が小さいことから，ICP と比べて姿勢推定に失敗することが少なく，安定した姿勢推定が可能であることがわかる．一方で t_z の並進誤差を比較すると ICP の誤差の方が小さい．これは，提案手法は距離情報を用いず推定しているためと考えられる．提案手法と 6D-PoseNet を比較すると，r_x の以外は提案手法の方が小さい．また，6D-PoseNet の分散を見ると，r_x を含めすべての誤差において提案手法の誤差の方が小さいことが確認できる．このことから提案手法は，反復処理を導入しているため 6D-PoseNet より安定した姿勢推定ができていると言える．

5. おわりに

本研究では，画像生成ネットワークを用いた反復処理による物体の姿勢推定が目標の姿勢に近似できることを確認した．今後の課題としては，物体のスケール変化への対応が考えられる．

参考文献

[1] P. J. Besl, *et al.*, “A Method for Registration of 3-D Shapes”, IEEE Trans. on PAMI, 1992.

[2] W. Kehl, *et al.*, “ SSD6D: Making rgb-based 3d detection and 6d pose estimation great again.” ICCV, 2017.

[3] V. Badrinarayanan, *et al.* “SegNet :A Deep Convolutional EncoderDecoder Architecture for Image Segmentation.”, IEEE trans. on PAMI, 2017.

研究業績

[1] 大西剛史 等, “面セグメンテーションに基づく 6D-PoseNet による位置姿勢推定”, 日本ロボット学会学術講演会, 2019.

[2] R. Araki, T. Onishi, *et al.*, “MT-DSSD: Deconvolutional Single Shot Detector Using Multi Task Learning for Object Detection, Segmentation, and Grasping Detection”, ICRA, 2020.