生活支援ロボットによる物体配置タスクにおける他物体との衝突予測に関する研究

TP20010 河合 竹彦　　指導教授：山下 隆義

1. はじめに

高齢者や身体障害者の生活を支援する介護者の人手不足が大きな社会問題になっている．そのため，家庭で家事や介護などの生活支援を行う生活支援ロボット [1] の導入が期待されている．生活支援において，コップ等の日用品を運ぶ物体運搬は重要な基本的タスクである．物体運搬タスクには，把持した物体を指定された場所に配置するタスクも含まれる．生活環境では様々な物体が乱雑に置かれていることが多いため，事前に他の物体と衝突が発生するかを予測し，衝突を防ぐ必要がある．物体配置タスクにおける衝突予測法として PonNet [2] が提案されている．PonNet は把持した物体を，乱雑に物体が置かれたエリア内に配置する際に衝突が発生する確率を予測する．入力情報はロボットが撮影した RGB-D 画像と把持した物体のモデル情報である．モデル情報は，既知の物体データベースを参照して得るため，未知物体を把持した場合に正しく衝突予測ができない．さらに，モデル情報のみでは現在の把持姿勢を考慮できず，把持姿勢の変化により発生する衝突を予測できない．

そこで本研究では，ハンドカメラを用いてアピアランスベースで把持した物体を捉える衝突予測法を提案する．アピアランスベースではモデル情報が不要で，把持した物体の形状や大きさ，把持姿勢等の特徴を捉えることができる．

2. 物体配置タスクの衝突予測

物体配置タスクの衝突予測手法である PonNet は，RGB と Depth の 2 つのモーダルの特徴を捉えるために，Attention Branch Network (ABN)[3] をマルチモーダルに拡張している．ABN を用いることで，特徴マップの重要な領域を強調しつつ，アテンションマップの可視化が可能である．人と生活支援ロボットの共同生活においては信頼性が重要である．PonNet はロボットの注視領域を可視化して判断根拠を明らかにすることで，信頼性の向上を図っている．また，生活環境のような乱雑に物体が置かれている状況下にて物体配置を行うと衝突が発生し，ロボットの破損や故障のリスクが高い．そこで，PonNet ではリアルな描画と連鎖的な物理作用を考慮できるシミュレータを用いて，衝突予測のためのデータセット BILA-12K を作成し，学習に用いている．

3. 提案手法

物体のモデル情報を用いる衝突予測法では，未知物体に対応できず, また，リアルタイムな把持姿勢情報を考慮できない．把持した物体をハンドカメラで撮影することで，アピアランスベースで物体の形状や大きさ，把持姿勢に関する特徴を捉えることができる．そこで，本研究では，ハンドカメラによって把持物体のアピアランスを捉えるマルチモーダル ABN を提案する．提案手法の流れを図 1 に示す．

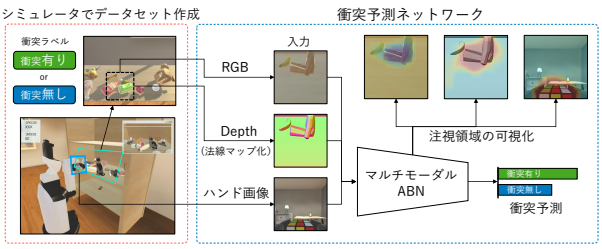


図 1: 提案手法の流れ

本研究では，生活環境を再現したシミュレータを用いて，把持姿勢の変化や未知物体を含む状況でのデータセットを作成する．このデータセットを用いてネットワークを学習し，衝突予測と注視領域の可視化を行う．

3.1 ネットワーク構造

本手法のネットワーク構造を図 2 に示す．各モーダル $k \in \{rgb, depth, hand\}$ に対応する 3 つの ABN，特徴に重

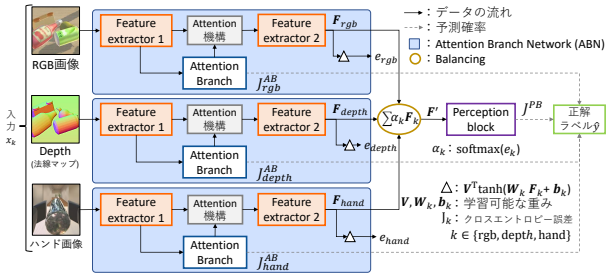


図 2: マルチモーダル ABN の構造

み付けする Balancing，判定を行う Perception block で構成される．ベースネットワークに ResNet-18 を使用し，第 3 レイヤーで分割して Feature extractor 1,2 に分ける．まず，各画像 x_k を Feature extractor 1 に入力して得られた特徴マップを Attention 機構と Attention Branch に入力する．Attention Branch の出力はクラス予測確率 $\boldsymbol{P}_k^{AB}(x_k)$ とアテンションマップである．Attention 機構によりアテンションマップで重み付けした特徴マップを Feature extractor 2 に入力し，256 次元の特徴ベクトル $\boldsymbol{F}_k$ を得る．これを，Blancing により重み付けして $\boldsymbol{F'}$ を得る．$\boldsymbol{F'}$ を Perception block に入力し，最終的な予測 $\boldsymbol{P}^{PB}(x)$ を得る．最終的な損失関数 J は式 (1) により求める．

$$J = J^{PB} + J^{AB}_{rgb} + J^{AB}_{depth} + J^{AB}_{hand} \tag{1}$$

このとき，各 J は, 各 $\boldsymbol{P}(x)$ と正解ラベル $\hat{y}$ のクロスエントロピー誤差で求める．

3.2 データセット作成

本研究では，把持物体のアピアランスを含む衝突予測データセットを，PonNet で構築されたシミュレータを用いて作成する．生活支援ロボットは，Human Support Robot (HSR) を対象とする．データセットの作成手順を Step1〜6 で示す．

Step 1　背景，照明環境と家具を選択．

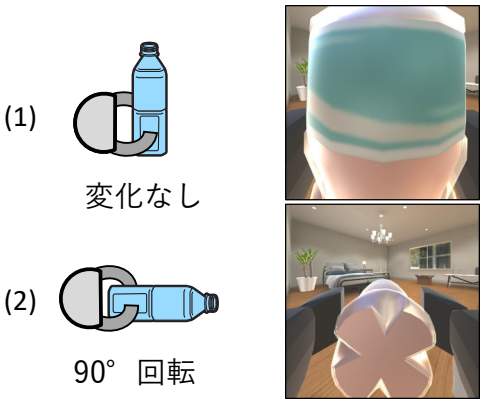
Step 2　28 種類の障害物の候補物体からランダムに選択し，配置エリア上に落下させ障害物として設置．

Step 3　配置エリアの状態を RGB-D 画像として HSR のヘッドカメラを用いて撮影．

Step 4　HSR が配置する物体を把持（把持姿勢は図 3(a)）．

Step 5　把持した物体をハンドカメラを用いて撮影 (図 3(b))．

Step 6　把持物体を配置エリアに配置し，衝突情報を記録．



(a) 把持姿勢　　(b) ハンドカメラ画像

図 3: 把持姿勢の種類

衝突有りの判定は障害物同士の衝突，ロボットアームと障害物の衝突，配置する物体と障害物の衝突，障害物と環境の衝突のそれぞれの条件で，閾値 0.1m/s 以上の衝突とする．これを必要なサンプル数が得られるまで繰り返し生成

し，データセットを構築する．実際に作成したデータセットの衝突ラベルの内訳を表 1 に示す．

表 1: データセットの衝突ラベル内訳 [%]

NDC	DC	O-O	T-O	A-O	O-D	P1	P2
69.67	30.33	3.15	21.11	14.41	10.46	23.19	76.06

NDC は衝突無しであり，DC は衝突有りである．DC は，下記の衝突条件のいずれか 1 つ以上が発生した場合である．O-O は配置エリア内の障害物同士の衝突を表す．T-O は配置した物体と障害物の衝突を表す．A-O は障害物とロボットアームの衝突を表す．O-D は障害物と配置エリアおよびその周辺環境への衝突を表す．P1，P2 は DC ラベルで把持姿勢 1，把持姿勢 2 の場合である．把持姿勢 2 の衝突有りラベルは把持姿勢 1 の 3.2 倍程度あり，把持姿勢 2 は衝突が発生しやすいことがわかる．このように，配置エリアの状態だけでなく，把持姿勢も考慮することが重要である．

4. 評価実験

本章では，提案手法の有効性を検証するための評価実験を行う．評価実験では物体配置タスクにおける衝突予測として NDC or DC の 2 クラス分類を行い，各手法との精度比較を行う．

4.1 実験条件

本実験には，前節で作成したデータセットを用い，学習に 42,100 回，検証に 5,300 回，評価に 5,300 回分のサンプルを使用する．ベースモデルには ResNet-18 を使用する．学習回数は 10 エポック，最適化関数に Adam を使用し，初期学習率を 10^{-5} とする．また，学習結果には，各学習エポックにおいて，最も検証精度が高いモデルの評価結果を用い，それぞれ 5 回学習した結果の平均を用いる．

4.2 既知物体に対する衝突予測の定量的評価

各モデルの衝突予測の正解率の比較を表 2 に示す．

表 2: 把持姿勢の変化有りの衝突予測の正解率

手法	モデル情報	ハンド画像	正解率 [%]
Ponnet[2]	✓	-	84.49 ± 0.22
	-	-	84.25 ± 0.30
提案手法	-	✓	**87.10 ± 0.33**

提案手法はハンドカメラの画像を利用し，精度が向上した．このことから，ハンドカメラの画像を利用することで，物体の把持姿勢や大きさの特徴を捉えていることがわかる．例えば，配置エリアの障害物の配置が同じであるが，物体の把持姿勢が違う場合に，ハンドカメラによって把持姿勢が得られ，適切な予測が可能になる．

4.3 アテンションマップの可視化による定性的評価

次に，正解ラベルと予測結果が DC の場合の正解例のアテンションマップの可視化結果を図 4 に示す．

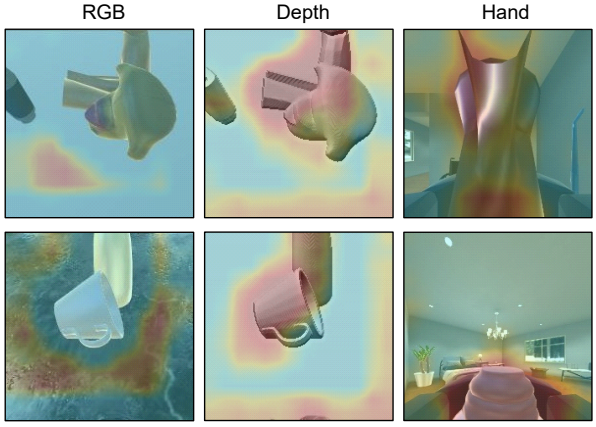


図 4: アテンションマップの可視化結果

図 4 の RGB 画像では，障害物が存在しない場所に注視領域が発生しており，平面を注視していることがわかる．また，Depth 画像では障害物に注視領域が発生しており，障害物を注視していることがわかる．このことから，RGB 画像と Depth 画像で異なる特徴を学習することで，精度が向上したと考えられる．次に，ハンド画像のアテンションマップの可視化結果では，把持した物体を注視していることがわかる．これによって，物体の種類や大きさと衝突の発生しやすさの関係などの特徴を学習していると考えられる．また，特定の物体のみにアテンションが発生する傾向あった．これは，把持姿勢 1 の状態や大きさが小さい物体など，衝突の危険性が低い物体とそうでない物体を区別していると考えられる．

4.4 Insertion によるアテンションマップの評価

アテンションマップの有効性の検証のために Insertion を行った結果を図 5 示す．

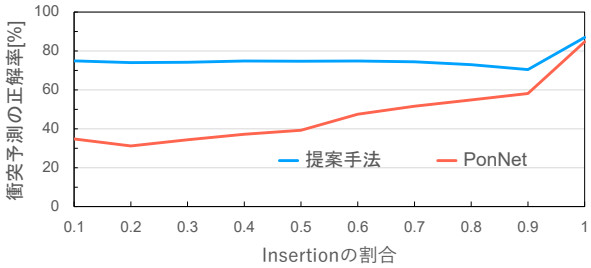


図 5: Insertion によるアテンションマップの評価

Insertion は入力画像に対するアテンションマップの各ピクセルの値を閾値以下を 0 にする処理を行い，入力画像をアテンションマップでマスクした後に，再度ネットワークに入力して推論を行う実験である．閾値を徐々に増加させた際の傾向から，アテンションが入力画像の重要な部分に発生しているか検証することができる．提案手法は，Insertion の割合が低い場合に精度が低下し，Insertion の割合が 1 のときに精度が最も高いことから，アテンションの値が高い部分が，入力画像の重要な部分を捉えていることがわかる．

4.5 未知物体に対する衝突予測の定量的評価

学習に使用していない未知物体に対する衝突予測を表 3 に示す．未知物体でも，PonNet と比較して提案手法では精度が向上しており，ハンド画像により物体の特徴を捉えていることがわかる．

表 3: 未知物体に対する衝突予測の正解率

手法	モデル情報	ハンド画像	正解率 [%]
PonNet	-	-	72.23
提案手法	-	✓	**74.06**

5. おわりに

本研究では，ハンドカメラによってアピアランスベースで把持物体を捉えるマルチモーダル ABN を提案した．また，把持物体のアピアランスを含む物体配置タスクの衝突予測のためのデータセットを作成した．作成したデータセットで学習した結果，ハンドカメラを追加したネットワークでの衝突予測の精度が向上した．今後は配置前の画像だけでなく配置後の画像を利用するなど，様々な条件での衝突予測と信頼性の向上を目指す．

参考文献

[1] T. Yamamoto. *et al.*: “Human Support Robot (HSR)”, ACM，2018.

[2] A. Magassouba. *et al.*: “Predicting and attending to damaging collisions for placing everyday objects in photo-realistic simulations”，AR，2021.

[3] H. Fukui. *et al*: “Attention Branch Network: Learning of Attention Mechanism for Visual Explanation”，CVPR，2019.

研究業績

[1] 河合竹彦，平川翼，山下隆義, 藤吉弘亘, 杉浦孔明,“配置物体を考慮した Attention 機構マルチモーダルネットワークによる衝突予測”, 日本ロボット学会学術講演会, 2021.

[2] 河合竹彦，平川翼，山下隆義, 藤吉弘亘,“把持動作結果を用いた追加学習による物体把持の高精度化”, 電気・電子・情報関係学会東海支部連合大会, 2020.