画像認識によるオブジェクト抽出とビジュアルフィードバックを用いたピッキングロボットシステム

Defective Article Picking Robot Using Image Processing Technique and Visual Feedback Control

◯ [1] 三木　康平,　[1] 永田　寅臣,　[2] 渡辺　桂吾
◯ [1]Kohei MIKI,　[1]Fusaomi NAGATA,　[2]Keigo WATANABE
[1] 山口東京理科大学大学院　[2] 岡山大学大学院
[1]Sanyo-Onoda City University　[2]Okayama University

Abstract: The authors are developing a robot system which can remove defective molded articles in narrow metallic mold space which has been manually done by skilled workers. The robot system can estimate the orientation of articles by using a transfer-learning based convolutional neural network (CNN). The orientation information is essential and indispensable to realize stable robotic picking. In addition, a visual feedback (VF) controller is designed by referring the COG position of articles obtained by image processing, so that the complicated calibration task between camera and robot coordinate systems can be eliminated. As a result, the authors propose a smart pick and place robot system which do not require conventional calibration tasks.

1　緒言

ロボットビジョン技術を用いたオブジェクトのピック & プレースは，長らく研究が進められてきた分野であり，Amazon Robotics Challenge といったピッキングを競う大会が開催されるなど現在も活発に研究が行われている．例えば，Yang らはオペレータが遠隔操作によりロボットを直接操作する感覚運動体験データを利用して，オブジェクトの折り畳みタスクに対し 77.8% の成功率を達成している [1]．また，小西らは，単眼カメラを用いて 3 次元物体の位置姿勢の認識を高速処理する手法を開発している [2]．ロボットビジョンを用いてピック & プレースを行う場合，カメラで撮影された入力画像からオブジェクトの物理空間上での位置情報および姿勢情報を推定する必要がある．

著者らは，前報にてオブジェクトの姿勢を 0° から 175° まで 5° ずつ，計 36 カテゴリの画像分類問題として，畳み込みニューラルネットワーク (CNN) モデルの一つである AlexNet を転移学習することにより，姿勢推定に特化した CNN を設計した．学習に使用していないテストデータを用いてこの転移学習ベースの CNN の汎可性を評価したところ，分類の認識率は 0.32 程度であり，±5° の誤差を許容した場合の認識率は 0.67 程度であった．その後，画像処理によりワークの重心位置を認識できるようにし，この CNN によりワークの姿勢を推定することでピック & プレースタスクを実行できるロボットシステムを提案した [3].

しかし，このシステムには次の二つの課題が存在する．一つは CNN による画像の認識率が低いことであり，もう一つは撮影するカメラとロボットとのキャリブレーション誤差のためにセットアップに時間を要することである．本論文では，このような課題を解決するために，VGG16 の転移学習による新たな CNN を設計し，分類性能の向上を図るとともに，ロボットコントローラにビジュアルフィードバック制御の機能を追加することで，作業テーブルを撮影するカメラとロボットとのキャリブレーションの工程の省力化を試みたので報告する．

2　CNN によるオブジェクトの姿勢推定

2.1　転移学習させる CNN

VGG16 は，13 の畳み込み層と 3 層の全結合層からなる CNN モデルであり，解像度 244×244×3 で受け取った画像を 1000 クラスのどれか一つに分類する．このモデルは，CNN の層の深さについて工夫されたきたネットワークの中で高い性能を出したモデルのうちの 1 つであり，モデル構造がシンプルであることや学習済みモデルが一般に公開されたことから，現在でも様々なモデルのベースネットワークや特徴抽出器として使用されている [5]．特に，ImageNet 大規模認識チャレンジ (ISVRC)2014 では，VGG16 と VGG19 を組み合わせた VGGNet がエラー率 7.3% で 2 位を獲得している（1 位は GoogleNet で 6.7%）．

転移学習とは，ある問題を効果的かつ効率的に解くために，別の問題での学習結果を再利用することであり [6]，その一つとして学習済みネットワーク (pre-trained network) を特徴抽出器として用いる手法がある．これは，重みを固定した pre-trained network を未学習ネットワークの上流に設置し，新しい学習用データを与えた際に pre-train network の適当な中間層からの出力をそのまま特徴ベクトルとして学習に用いるものであり，深層学習や CNN に関する専門知識がなくとも手軽に新しい CNN を作成することができる [7].

今回，学習させる VGG16 ベースの CNN は，図 1 に示すように VGG16 の最後の全結合層 (Fully connected layer: FC 層) および最終分類層を未学習の層に置き換え

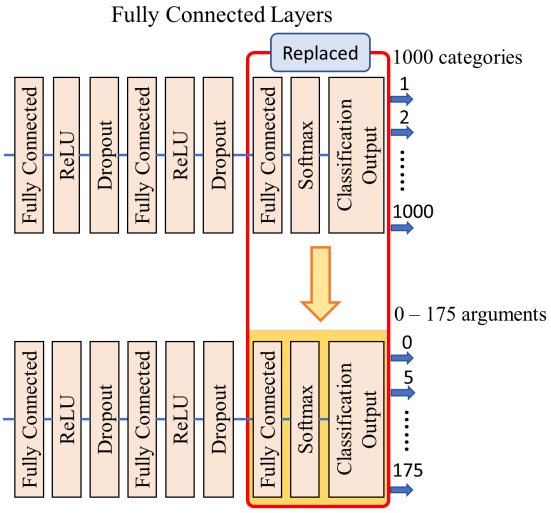

Fig. 1: Replacement of fully-connected layers of VGG16 to cope with orientation detection.

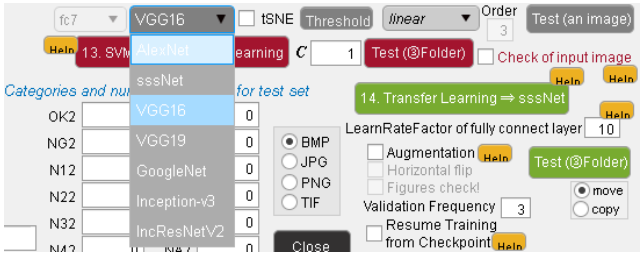

Fig. 2: A part of the developed design tool for transfer learning based CNNs .

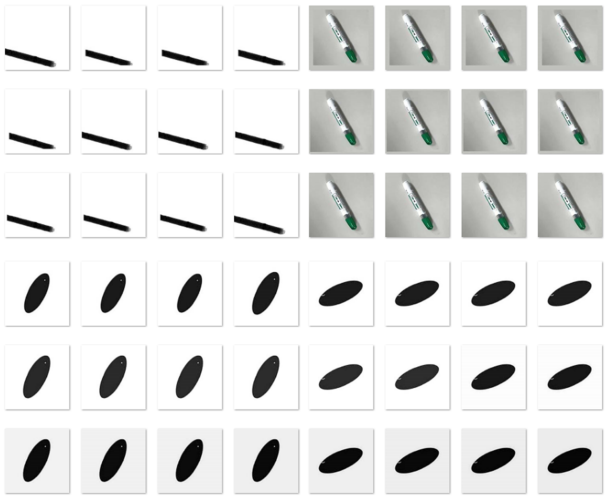

Fig. 3: Example of training images including four different angles, i.e., 15°, 60°, 115°, 155°.

たものである.

2.2 CNN の学習

CNN の学習には，現在開発している CNN とサポートベクタマシン (SVM) 設計＆訓練ツール [3] を用いた．図 2 には，この設計ツールのダイアログ内の転移学習設定部分を示す.

学習は，最大エポック数: 500，ミニバッチサイズ: 30，目標の認識率: 0.999，目標誤差: 0.0001，L_2 正則化係数: 0.004 を設定し，最適化関数にモーメンタム項付き確率的勾配降下法を使用して行った．このほか，CNN の過学習を防止するため，ミニバッチサイズ 1 回分の訓練が終了するたびに各カテゴリ 99 枚の検証用画像を用いて汎化性能が低下していないかを検証した．学習の終了条件は，ミニバッチサイズ単位で連続して 200 回の訓練期間で検証データに対する認識率が一度も向上しなかった場合に停止するとした．この値を小さく設定すると，誤差が収束する前に学習が終了してしまうため，今回は 200 を設定した．CNN の学習には，NVIDIA GeForce GTX 1060 6GB を搭載したコンピュータにて 24 時間を要した.

学習用画像には，これまでに蓄積してきた 36 カテゴリに分類された計 15,264 枚の画像 (1 カテゴリあたり 424 枚) を用いた．図 3 には実際に使用した画像のサンプルを示す．この画像は，299×299 ピクセルのグレースケール画像および RGB 画像で構成されており，CNN&SVM 設計ツールの画像オーギュメンテーション機能を用いて，オリジナル画像に対し，明暗の変化，画像の拡大縮小，±1° の角度変化を与え作成した．訓練，検証およびテストに用いた画像はすべて，正方形に切り取ったものを使用した．これは，VGG16 への入力画像のサイズが 224×224×3 と固定されているため，アスペクト比のばらつきによる分類認識率の低下を防ぐためである.

2.3 姿勢の推定結果

同じデータセットを用いて VGG16 ベースの転移学習を 3 回行い，それぞれモデル番号 No.1, 2, 3 として保存した．同様に，AlexNet ベースの転移学習を 3 回行い，それぞれを姿勢推定用のモデル No.4, 5, 6 として保存した．その後，図 4 に示すような学習時とは大きく異なる特徴を持つ未学習のテスト画像 (1 カテゴリあたり 57 枚) を与え，分類させることでこれらのモデルを比較評価した．その結果を表 1 に示し，モデルごとの画像のクラスと誤分類枚数の関係を図 5 に示す．誤分類枚数は, それぞれの 3 つのモデルの平均値を示している．表 1 内の下段の Accuracy の値は ±5° の誤差を許容した場合の認識率である.

Fig. 4: Examples of test images to compare the generalization ability between VGG16-based and AlexNet-based transfered CNNs.

Table 1: Comparison of classification results between two CNNs transferred based on AlexNet and VGG16. The lower row shows the accuracy with ±5°tolerance.

Model	VGG16-based			AlexNet-based		
No.	1	2	3	4	5	6
Accuracy	0.55	0.49	0.48	0.25	0.24	0.24
Accuracy	0.84	0.77	0.81	0.61	0.59	0.58

画像分類による姿勢推定の認識率は，VGG16 ベースのモデルの方が AlexNet ベースのものよりも 25% 程度高く，VGG16 を使用することの有効性を確認することができた．このとき，VGG16 は 35°, 85°, 115°, 145° の 4 つで分類の認識率が高く，AlexNet は 15°, 65° の認識率で VGG16 を超えていた．また，15° から 25° と 140° から 155° の部分では，どちらもモデルも誤分類枚数が少なくなっていた．

3 実機ロボットを用いたピック&プレース実験

3.1 システムの構成と画像処理

図 6 にはピック&プレースロボットの構成を示す．ロボット本体には Dobot 社製の小型ロボットアームである Dobot Magician を用い，アーム先端にはグリッパタイプのエンドエフェクタを備えている．ピック&プレース実験は，図 7 に示すような手順を経て行う．まず，カメラでテーブル上を 1280×960 ピクセルの領域として撮影し，オリジナル画像とする．次に，その画像をグレー階調化後

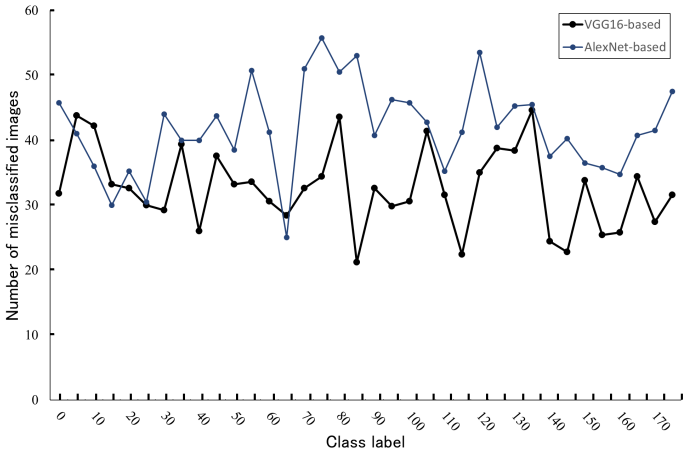

Fig. 5: Relationship between class labels and number of misclassified images in two CNNs.

に二値化処理し，最大面積を持つ連結成分をオブジェクトとして検出する．そして，画像座標系におけるオブジェクトの重心位置 $\boldsymbol{I} = [I_x\ I_y]^T (1 \leq I_x \leq 1200, 1 \leq I_y \leq 960)$ は，ピクセル集合と等面積で厚みが一定，かつ慣性モーメントが等しい楕円 (相当楕円) の重心として，画素値 P とピクセル座標 (p, q) より次式を用いて求めることができる．

$$I_x = \frac{\sum_{p=1}^{x}\sum_{q=1}^{y} xP(p,q)}{\sum_{p=1}^{x}\sum_{q=1}^{y} P(p,q)}$$
$$I_y = \frac{\sum_{p=1}^{x}\sum_{q=1}^{y} yP(p,q)}{\sum_{p=1}^{x}\sum_{q=1}^{y} P(p,q)} \quad (1)$$

ここで, 画像の原点は左上とする．さらに，$\boldsymbol{I}$ をロボット座標系におけるオブジェクトの重心位置 $\boldsymbol{G} = [G_x\ G_y]^T$ に変換するために次式を用いた．

$$G_x = X_1 + I_x \frac{X_2 - X_1}{1200} \quad (2)$$
$$G_y = Y_1 + I_y \frac{Y_2 - Y_1}{960} \quad (3)$$

ここで，(X_1, Y_1) と (X_2, Y_2) はそれぞれ，画像の左上と右下に対応したロボット座標系における位置である．重心位置 $\boldsymbol{I}$ の推定後は，検出したオブジェクトの相当楕円の長軸の長さを基準として，オリジナル画像上のオブジェクト周辺を長軸の 110% のサイズで正方形に切り取る．この画像を今回設計した CNN の入力として与えることで，オブジェクトの姿勢を 0° から 175° の範囲で 5° 刻みでラベル化した 36 カテゴリのどれかに分類し，姿勢推定を行う．このとき，ラベルとなる姿勢角度 θ の基準は, 画像の下辺とした．

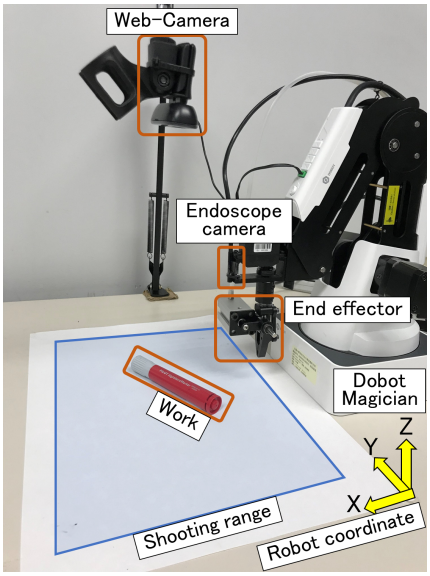

Fig. 6: Pick and place robot incorporated with the CNN transfer-learned with VGG16.

3.2 Visual Feedback 制御

Visual feedback(VF) 制御は，図 6 に示すように Dobot のエンドエフェクタ付近に新たに設置した小型の内視鏡カメラを用いて行う．本制御の目標は，カメラで撮影した画像の中心座標とオブジェクトの重心位置座標との偏差が 0 になるようエンドエフェクタの位置を制御することであり，その偏差は次式で計算される．

$$\boldsymbol{e}(k) = \boldsymbol{X}_d - \boldsymbol{I}(k) \tag{4}$$

ここで，k は離散時刻であり，$\boldsymbol{X}_d = \left[\frac{X_r}{2}, \frac{X_r}{2}\right]^T$ はカメラ画像の中心座標である．この偏差 $\boldsymbol{e}(k)$ をもとに，エンドエフェクタの動作速度 $\boldsymbol{v}(k) = [v_x(k), v_y(k)]$ を操作量とし，次式により PI 制御を適用する．

$$\boldsymbol{v}(k) = K_p \boldsymbol{e}(k) + K_i \sum_{n=1}^{k} \boldsymbol{e}(n) \tag{5}$$

ここで，K_p と K_i はそれぞれ P 制御と I 制御のゲインである．これにより画像の中心に内視鏡の位置を制御することができるようになった．なお，エンドエフェクタの先端をその位置まで移動させるには，X 軸方向に内視鏡の中心からエンドエフェクタの中心位置までのオフセット分だけ移動させればよい．このオフセット値を実測した結果，38.925mm であった．

3.3 評価用オブジェクトのピック & プレース実験

画像処理によるオブジェクト重心位置の検出機能，VF 制御によるワークへのアプローチ機能，VGG16 の転移学習ベースの CNN によるワークの姿勢推定機能により，キャリブレーションを行うことなく作業テーブル上の任意の位置に任意の姿勢で配置した USB メモリやペンなどのテストワークを良好にピック & プレースさせることができた．図 8 には実験風景を示す．

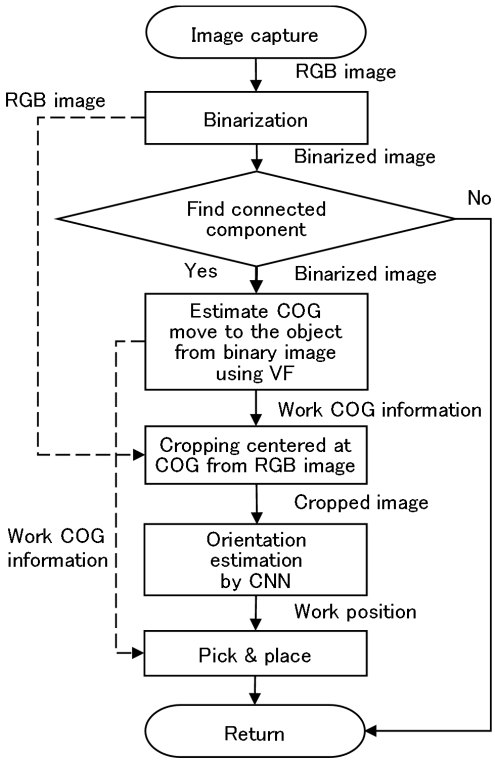

Fig. 7: Flow chart of the robotic pick and place operation, in which the proposed CNN and VF control are included.

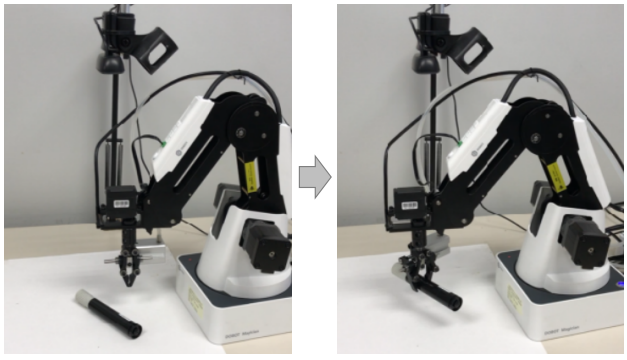

Fig. 8: Robotic pick and place experiment.

4 結言

本研究では，転移学習のベースとなる CNN モデルを AlexNet から VGG16 に変更することによって姿勢の認識率を向上させることができた．また，VF 制御を追加することにより，手動キャリブレーションの基準座標設定時に発生する位置決め誤差を逐一修正する必要がなくなり，ピック & プレースタスクを容易に実行できるようになった．今後の課題として，ロボットの姿勢によっては，VF 制御用カメラで撮影した画像内にエンドエフェクタが写り込み，それによって動作が停止することがあったため，カメラの固定位置について再検討していく．また，カメラが装着されたロボットアーム先端を挿入できないような狭い空間にある不良品ワークの検出と姿勢推定ができるように，斜め方向から撮影したワークの画像を訓練データに用いた CNN を構築していく予定である．

参考文献

[1] P.C. Yang, K. Sasaki, Kanata Suzuki, Kei Kase, Shigeki Sugano, Tetsuya Ogata, “ Repeatable Folding Task by Humanoid Robot Worker Using Deep Learning,” *IEEE Robotics & Automation Letters*, Vol. 2, No. 2, pp. 397–403, 2017.

[2] 小西 嘉典, 半沢 雄希, 川出 雅人, 橋本 学, “階層的姿勢探索木を用いた単眼カメラからの高速3次元物体位置姿勢認識”, 電子情報通信学会論文誌 D, Vol. J100–D, No. 8, pp. 711–723, 2017.

[3] 三木 康平, 永田 寅臣, 渡辺 桂吾, “金型空間内の不良成型品ピッキングのための画像認識を用いたロボットシステムの検討”, ロボティクスメカトロニクス講演会講演予稿集 2020, 2P2-B03, 2020.

[4] F. Nagata, K. Miki, Y. Imahashi, K. Nakashima, K. Tokuno, A. Otsuka, K. Watanabe and M.K. Habib, “Orientation Detection Using a CNN Designed by Transfer Learning of AlexNet,” *Procs. of the 8th IIAE International Conference on Industrial Application Engineering 2020*, pp. 295–299, 2020.

[5] Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks For Large-ScaleImageRecognition,” *Procs. of International Conference on Learning Representations 2015 (ICLR2015)*, pp. 1–14, 2015.

[6] 神嶌 敏弘, “転移学習”, 人工知能学会誌, Vol. 25, No. 4, pp. 572–580, 2010.

[7] 中山 英樹, “深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習”, 電子情報通信学会技術研究報告, Vol. 115, No. 146, pp. 55–59, 2015.