

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

Defective Article Picking Robot in Narrow Metal Mold Space Using Image Processing Technique

○学 三木 康平（山口東京理科大学大学院） 正 永田 寅臣（山口東京理科大学大学院）
正 渡辺 桂吾（岡山大学大学院）

Kohei MIKI, Graduate School of Science and Engineering, Sanyo-Onoda City University
Fusaomi NAGATA, Graduate School of Science and Engineering, Sanyo-Onoda City University
Keigo WATANABE, Graduate School of Science and Technology, Okayama University
f116056@ed.socu.ac.jp, nagata@rs.socu.ac.jp, watanabe@sys.okayama-u.ac.jp

In this paper, we report the result on basic research using a pick and place robot for removing resin mold articles that have unfortunately fallen on one of molds in an injection molding machine in which multiple molds are arranged for mass production. The position and orientation of target workpieces are essential for successfully conducting a picking task, so that they are tried to be estimated using our developed software for image processing and design of convolutional neural network (CNN). The CNN is one of typical AI technologies specialized in image recognition. The CNN used in experiments is designed through AlexNet-based transfer learning, for estimating the orientations of workpieces. The effectiveness of this system is evaluated through pick and place experiments using an articulated robot named DobotMagician.

Key Words: Convolutional neural network, Transfer learning, Picking robot

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&ブレースは、長らく研究が進められている分野であり、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% の成功率を達成した [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た [2]。本研究では射出成型機内に複数配置された金型上に落下した不良成型品のピッキング作業に着目する。この作業を自動化するには 2 つの課題が存在する。一つは、金型表面の凹凸や不均一な光沢面の影響下でも、ワークの位置と姿勢を正確に推定するための画像処理に関する課題、もう一つは射出成型機で使用される金型空間内に落下した不良成型品検出のために、金型全景を真上から撮影できる位置にカメラを挿入するための十分なスペースがないというシステムの構造に関する課題である。これらの課題を解決するために、まず不良品がどの金型上に存在するかを射出成型機内に配置した複数金型の全景撮影用のカメラで大まかな位置を認識し、その後、小型カメラを搭載した金型空間内に挿入可能なロボットアームをビジュアルフィードバックによってターゲットの不良品上へアプローチさせ、ピック&ブレースを行うシステムを検討している。本稿では、これまでに開発してきたピッキング自動化のための基本システムについて報告する。

2 CNN によるオブジェクトの姿勢推定

ロボットアームを用いてオブジェクトをピッキングし整列させる動作や、他のものに組付けるといった動作を行う場合には、ピッキング前のオブジェクトの位置と姿勢情報を正確に把握しておくことが重要である。このため、現在開発している畳み込みニューラルネットワーク (CNN) とサポートベクタマシン (SVM) の設計と訓練ツール [3] を用いて、AlexNet の転移学習によりオブジェクトの姿勢推定に特化した CNN を設計した。以下、この CNN について説明する。

2.1 CNN 設計 & 訓練ツール

図 1 には、今回使用した CNN & SVM 設計ツールのダイアログログを示す。開発は Matlab 上で行い、NeurlNetwork Toolbox, Parallel Computing Toolbox, Deep Learning Toolbox などのパッケージを用いている。このツールには例えば、次のような機能がある。

- 転移学習による新たな CNN モデルの設計
- CNN を特徴抽出器として用いた SVM 分類器の設計
- 複数の CNN と SVM を組み合わせたアンサンブル学習モデルの設計

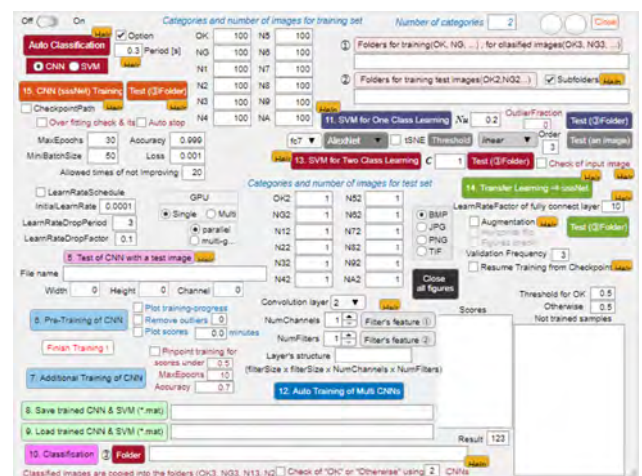


Fig.1 Developed Design tool for CNN and SVM.

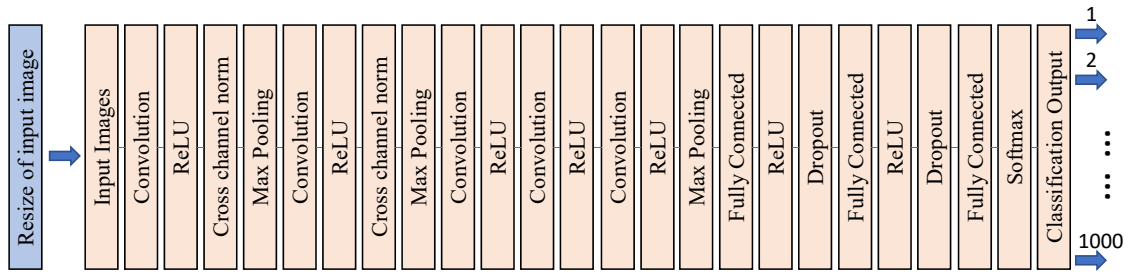


Fig.2 Structure of original AlexNet for classification of 1000 categories.

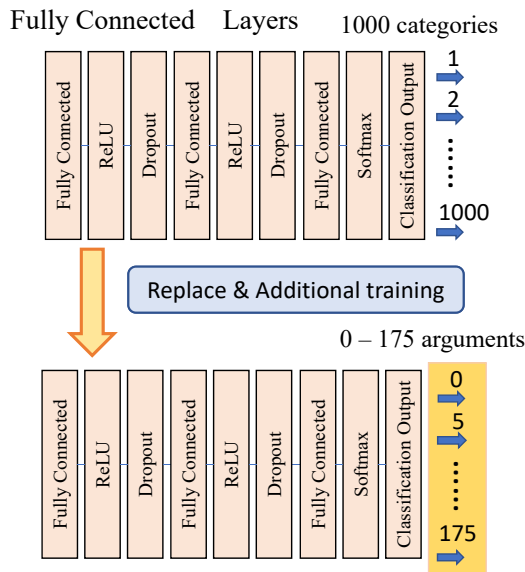


Fig.3 Replace of fully-connected layers of AlexNet.

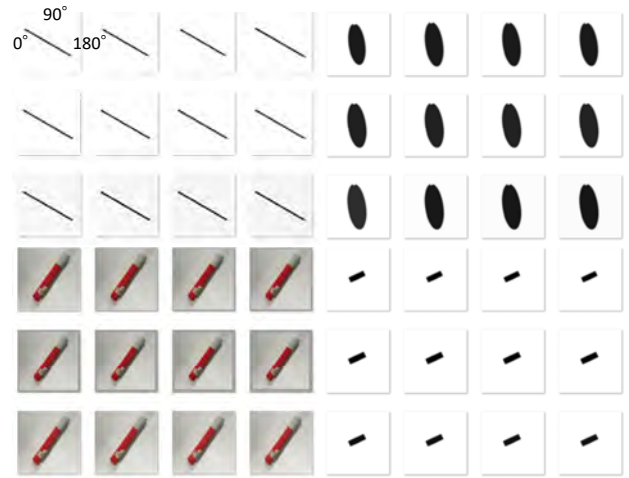


Fig.4 Example of training images including four kinds of angles.

2.2 AlexNet の転移学習による姿勢推定用 CNN の設計

画像認識に特化した CNN では、畳み込み層がカテゴリ毎の画像に共通した特徴を学習し、全結合層がそれらの特徴と最終的に分類されるカテゴリとの関係を学習することが経験的にわかっている。転移学習とは、大規模データセットで事前に学習した既存の優れた CNN モデルをもとに、ある領域に特化したモデルを新しい訓練データで再構築する手法であり、学習時間の短縮やデータ数が少ない場合でも効果的な CNN モデルを構築できるといった特徴がある。そこで、既存の CNN モデルとして 1,000 カテゴリの分類が可能な AlexNet を選択し、その全結合層部分を図 3 のように 0 から 175 までも 5 刻みに分けた 36 カテゴリの分類が可能な未学習の構造に置き換え、準備した訓練データを用いて追加学習を行うことでオブジェクトの姿勢推定用に再構築する。AlexNet とは、図 2 のような構造を持つ優れた CNN モデルの一つであり、ImageNet LSVRC2012 という画像認識コンペティションにて優勝した実績を持つ。このモデルは、鳥や飛行機、車や犬など 1,000 カテゴリに分類された合計 120 万枚程度の画像で訓練されており、画像に対して優れた特徴抽出能力を獲得している。そのため、入力画像はそこに含まれる特徴から 1,000 カテゴリの内の一つに高い精度で分類される。

2.3 学習と推定結果

追加学習には、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 4 には訓練に用いた 25°, 80°, 125°, 140°, の姿勢情報を含む画像のサンプルを示す。この訓練用画像は、299×299 ピクセルのグレースケール画像およ

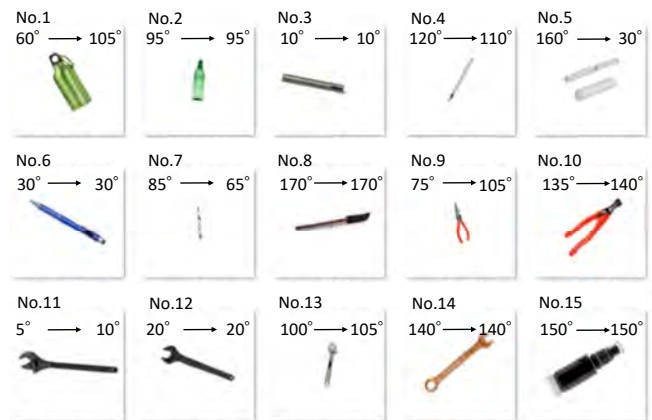


Fig.5 Test images used to evaluate the CNN generalization.

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、 $\pm 1^\circ$ の角度変化を与えて作成した。また、転移学習の主なパラメータは最大エポック数: 500、ミニバッチサイズ: 100、目標精度: 0.999、目標誤差: 0.001 に設定した。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確認を行いながら学習をすすめた。学習の終了条件は、連続したミニバッチサイズ 50 回分のデータで訓練中、検証精度が一度も向上しなかった場合に停止するとした。このような学習プロセスの結果、最終的に得られた検証精度は 98.93% であった。CNN の学習には NVIDIA GeForce GTX 1060 6GB を搭載したコンピュータにて 2 時間を要した。

つぎに、転移学習によって得られた CNN に対して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 15 枚を与え、分類させることでその汎化性能を確認した。なお、今回の CNN の訓練、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、AlexNet への入力画像のサイズが $229 \times 229 \times 3$ 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精度の低下を防ぐためである。図 5 には、テストに使用した画像および、その左側に画像のラベルとなる角度を、右側には CNN による推定結果を示す。この分類テストではデータセットに含まれていなかった形状を持つオブジェクトに対して、およそ 5 割の確率でしか 5 度刻みの厳密な分類を行うことができていなかったが、ピッキングの成否に影響ないと思われる 10 度程度の誤差を許容すると 7 割以上の正解率であった。許容できる範囲で姿勢を推定できなかった画像は Nos. 1, 5, 9 であったが、例えば、No. 5 の画像については複数のオブジェクトが存在しており、今回の CNN では特徴を抽出できなかったものと考えられる。しかし、次章の実験では複数ワークは対象としないため、ロボットシステムへの実装は問題ないと判断した。

3 実機ロボットを用いたピック&ブレース実験

本章では、前章で設計した姿勢推定用の CNN を実装した小型アームロボットを用いてピック&ブレース実験を行い、有効性を検証する。

3.1 システムの構成と目標とする動作

図 6 にはピック&ブ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 を使い、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。このほか、作業テーブル上を撮影する WEB カメラ、ピッキング対象のワーク、MATLAB が動作する制御用 PC で構成されている。図 7 には、MATLAB の App Designer 上で開発したこのシステムの制御用ダイアログを示す。ピック&ブレース実験は、図 8 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、その画像をグレースケール化後に二値化処理し、最大面積を持つ連結成分を対象オブジェクトとして検出し、画像座標系におけるオブジェクトの重心位置 $[X_x \ I_y]^T$ ($1 \leq X_x \leq 1200$, $1 \leq I_y \leq 960$) を計算する。さらに、 $[X_x \ I_y]^T$ をロボット座標系におけるオブジェクトの重心位置 $[G_x \ G_y]^T$ に変換するために次式を用いた。

$$G_x = X_1 + I_x \frac{X_2 - X_1}{1200} \quad (1)$$

$$G_y = Y_1 + I_y \frac{Y_2 - Y_1}{960} \quad (2)$$

ここで、図 9 に示すように (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 $[X_x \ I_y]^T$ 検出後は、検出した対象オブジェクトと等面積で、かつ慣性モーメントが等しい楕円の長軸の長さを基準として、オリジナル画像上の対象物周辺を長軸の 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 度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、図 9 のように定めた。

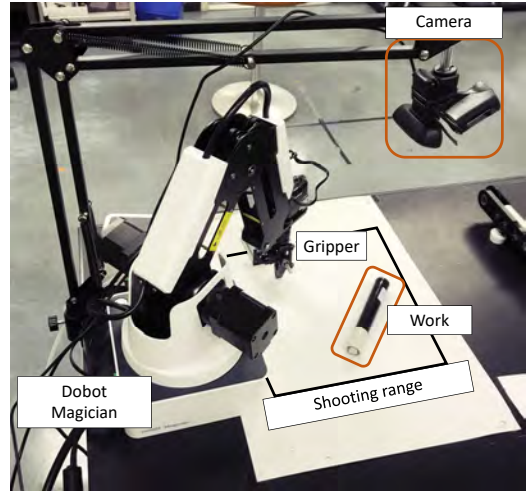


Fig.6 Pick and place robot incorporated with the CNN transfer-learned with AlexNet.

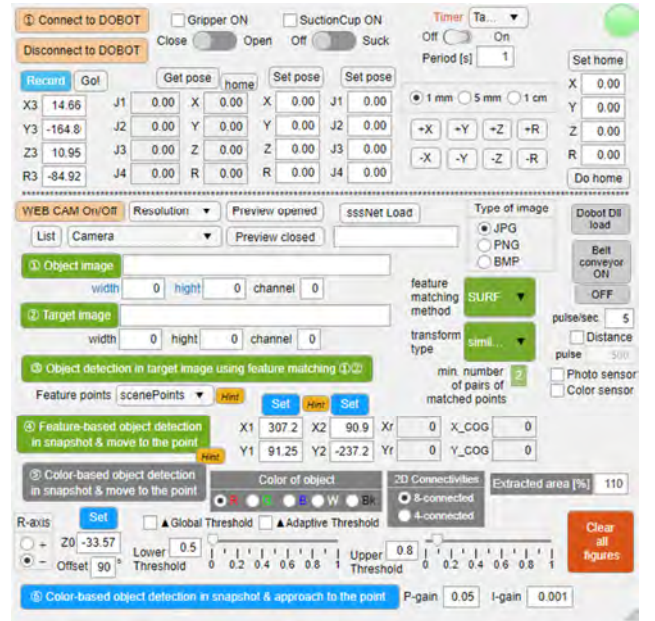


Fig.7 Control dialogue developed for the pick and place robot.

実験では、この推定された姿勢情報をもとに、オブジェクトとグリッパの開閉方向が直交するような最適角度にグリッパを回転させることで、ワークを良好に把持することができるものと期待される。

3.2 オブジェクトのピック&ブレース実験

ピック&ブレース実験で対象とするワークには、図 10 に示すような訓練画像に含まれていなかった 6 つのワークを選んだ。これらは、ワークが姿勢情報としての特徴をもつことと、ある程度透明度を持った樹脂成型品であることなどを理由に選択した。ピック&ブレースの実験は、それぞれのオブジェクトを任意に変化させて 10 回ずつ試行し、把持の成否を記録した。その結果を表 11 に示す。この結果より、どのワークに対しても高い精度でピッキングが行えたことが確認できる。特に、Nos. 1, 2, 3 のように色が

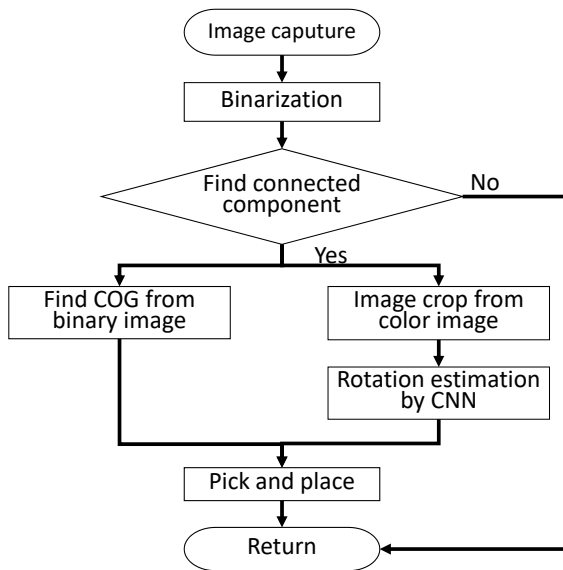


Fig.8 Flow chart of pick and place operation.

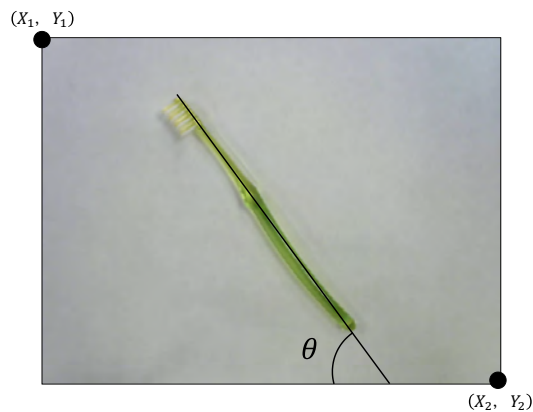


Fig.9 Relation between captured image and robot coordinate system.

濃いワークほど、ピッキング精度が高かった。これは、背景とのコントラストが高いほど、二値化後に行う対象オブジェクトの重心位置の推定の精度が向上することが原因であると考えられる。一方、No. 4, No. 5 のように表面に光沢をもつワークや、透明度の高いワークは、光の反射の状況によって画像処理の性能が安定しないため、その情報をもとに行う把持動作に影響が出たものと考えられる。

4 結言

本研究では、転移学習をベースに設計したオブジェクトの姿勢推定に特化した CNN を搭載したピックアンドブレースロボットを提案した。未学習のワークのピック&ブレース実験により、提案システムの有効性が確認された。今回は、WEB カメラで撮影したシーン画像内にワークが存在することが前提となっており適用が比較的容易であった。課題としては、ワークの透明度が増すと二値化処理において背景とワークを分離することが困難となる傾向があったため、このようなワークに対しての位置認識・姿勢推定精度を高めるために訓練データを増やししながら CNN を改善していく。

今後は、射出成型機内に並んだ複数の金型内のどこに対象とする小さな樹脂成型品ワークの不良品が落下し、存在しているのか



Fig.10 Test objects to be picked.

Fig.11 Number of successful results of pick and place.

No.1	No.2	No.3	No.4	No.5	No.6
10	9	10	8	8	9

をまず検知し、その後、ビジュアルフィードバック制御によりグリッパをワーク上部へ移動させることで画像座標系とロボット座標系のキャリブレーションを行うことなくピッキング動作を実行できるシステムの実現を目指していく。

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 AND 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] 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*, 5 pages, Kunibiki Messe, Matsue, Shimane, Japan, March 26–30, 2020.