

調理行動の理解に向けた 切断作業における食材と加工方法の同時識別

信州大学大学院 総合理工学研究科 工学専攻 ○池神奈穂美, 藤波香絵, 山崎公俊

本稿では, 調理における切る加工において, 加工対象となる野菜の種別と切り方を同時に判別するシステムの構築を目的とし, 動作の認識で主に用いられるカメラ画像から, CNN を用いた画像識別により実現させる. 本研究では, 2つの種類を同時に識別するにあたり畳み込み層のみ重みパラメータを共有し, 全結合層を識別種類ごとに独立させることを提案する. 提案したシステムは検証するための実験をし, 同じネットワーク構造による単独種類での識別精度には劣るが, 87.9%精度で同時に識別可能であることを確認したので報告する.

キーワード: 調理, 動作識別, CNN, 画像識別

1. 緒言

近年, 様々な作業の機械化および自動化が進んでいる. その範囲は, 生産の場である工場だけでなく, 家庭の場にも及び, 調理もその一つである. 調理の機械化に関しては, 炊飯器や製麺機など特定の作業を代替する製品が存在するが, そのどれもが人手によるサポートが必要とされる. 一方で, 人型ロボットなどにより人と同じように調理させる研究が進んでおり, そのための研究の第一歩として人の調理行動を認識する研究が多く行われている. Amaro ら[1]による研究では, 色判別を用いて人の調理行動を認識, 識別している. 齋ら[2]は料理番組映像から「切る・炒める・混ぜる」といった基本的な動作を繰り返す特徴から調理動作の抽出を行った.

調理をする上で必要となる情報は, 食材に対してどのような道具でどのような加工をするかである. 特に, 食材によって加工法は異なるため, 食材と加工法を同時に判断する必要がある. Yamakata ら[3]は, 調理における準備段階としての皮むきおよび切る作業の認識と調理中の手順全体の認識のように, 作業を二つに分けた食品の認識手法を提案したが, 本研究では, 調理動作のうちの切る動作のみに焦点を当て, 食材の識別と同時にを行うことを目的とする.

2. 同時識別手法

2.1 同時識別のアプローチ

調理に限らず, 動作認識の多くにはカラーカメラが用いられる. 本研究においてもカメラから得られるカラー画像から識別する方法を考える. 画像をもとに識別する方法として, 畳み込みニューラルネットワーク (Convolution Neural Networks, CNN) がある. これは, 多次元・多層入力に対し, 畳み込みフィルタを介して次元圧縮をおこない, 識別結果を得る手法である. 例として手書き文字の識別が挙げられる.

2.2 画像による同時識別手法の提案

畳み込み層と全結合層で構成される一般的な画像識別ネットワークでは, 1種類のクラス分類しか行わず, 複数種類の同時クラス分類はできない. しかしながら, 畳み込み層によって得られる入力画像に対する特徴量が同じであっても, その後段に位置する全結合層により, 複数種類の識別結果を得ることは可能である. そこで, 次のような手順で, 同時識別

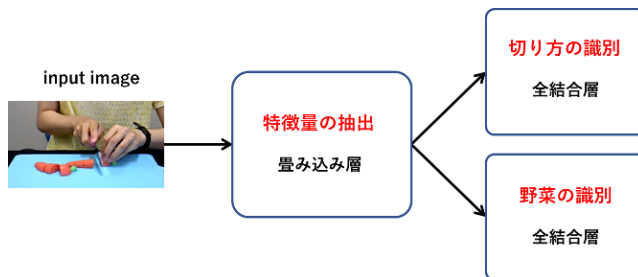


図 1 同時識別のためのシステム構成図

可能なネットワーク構造を構築する.

- 複数種類の識別対象それぞれについて, 1種類のクラス分類を行う画像識別ネットワークを構築し, 高精度な識別を可能とする構造を探す.
- 複数種類のクラス分類を同時に行うため, 上記の画像識別ネットワークをもとに全結合層を種類ごとに独立化させる.

図1にネットワーク構造の概要を示す. 入力画像および畳み込み層は共有し, その後段にある全結合層のみ, 識別の種類数に応じて用意する. 誤差関数および学習アルゴリズムを次のように設定する.

- 識別種類ごとに次の誤差関数 $loss_{type}$ を使用する.

$$loss_{type} = \sum -\log P_{type}$$

- 学習アルゴリズムにAdamOptimizerを使用し, 識別種類ごとの誤差 $loss_{type}$ の合計値を最小化するようにパラメータを変化させる.

3. 画像識別実験

3.1 ネットワークの設定

ネットワークの構造は, 事前検討の結果次のようにした. 畳み込み層の層数: $32 \times 32 \times 64$ の3層, カーネルサイズ: 1層目において 1×1 , その他の層において 3×3 , ストライド: 1×1 . また, zero padding および max pooling 処理を加えた. 入力はRGB画像の 128×256 の3チャンネルとした. この設定のもとでは, 2層の全結合層を通した出力は1024次元となった.

3.2 学習データセット

機械学習には総じて大量のデータが必要とされる。本研究では食材を切る作業が識別対象であるが、データを収集するうえで、1度切ってしまった食材を使いまわすことは難しい。また、食材の経時劣化の問題もある。よって、労力と費用の両面で大量のデータ収集は困難であると言える。そのため、木材等で造形された模擬食材を用いた。

高さ 70.0cm のテーブル上にまな板を設置し、椅子に座って切断作業を行った。このとき、テーブルの端より 79.5cm 離して、高さ 0.9m に、0.27rad 傾けて 3 次元距離センサ Kinect を設置した。画像は作業中の手元およびまな板全体が写るように 150×270 のサイズに切り取り、保存した。図 2 にデータセットの一例を示す。



図 2 切り方のデータセット例

本研究では、少ないデータで学習を行うために、学習時のバッチ作成の際に、ランダムクロッピング処理により 128×256 のサイズへ切り出し、コントラスト変換処理を行うことで、データの水増しした。また、収集された 246 枚または 327 枚の画像データは、train データに 80%、その他を test データに使用し、バッチ数 25 で学習させた。

3.3 切り方および野菜の種類におけるクラス分類

はじめに 1 種類の画像識別のために構築したネットワークについて、識別精度の検証を行った。食材の切り方については、乱切り、輪切りおよび半月切りの 3 クラスに、野菜の種類については人参、大根およびじゃがいもの 3 クラスに分類した。表 1 にそれぞれの画像識別ネットワークを 2000step まで学習させたときの誤差と精度の結果を示す。

表 1 各種類の学習誤差と識別精度

	loss	Accuracy
切り方	2.849	96%
野菜	2.938	96%

表 1 より、切り方および野菜の種類の識別、どちらにおいても 96% と高い精度の識別結果を得られた。したがって、本実験で用いたネットワーク構造は適切に機能したといえる。

3.4 切り方と野菜の同時識別によるクラス分類

前節の実験のネットワーク構造に対し、全結合層を 2 つそれぞれ独立に用意し、切り方と野菜の同時識別によるクラス分類精度の検証実験をした。表 2 に画像識別ネットワークを 2000step まで学習させたときの誤差と精度の結果を、図 3 に学習に伴う誤差の変化を示す。

表 2 同時識別による各種類の学習誤差と識別精度

	loss	Accuracy
切り方	57.701	87.9%
野菜	6.068	99.8%

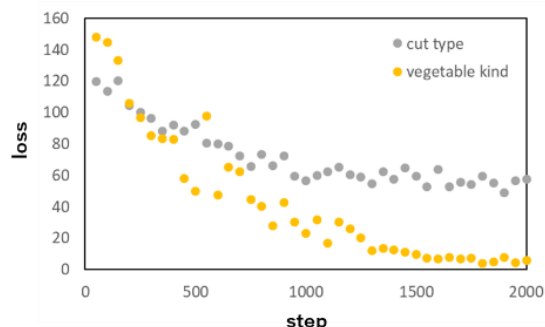


図 3 学習に伴う誤差の変化

表 2 より、切り方に対し 87.9%、野菜の種類に対し 99.8% の精度で識別できた。前節の単独種類における識別結果に比べ、切り方の識別精度は下がったが、精度としては高い水準を保つことができた。識別精度に差が生じた原因としては、誤差の最小化方法が挙げられる。本実験では、それぞれの識別種類ごとに誤差の合計値を最小化させるように学習させた。これに対し、誤差についても種類ごとに独立して最小化させるように学習させることで改善できると考えられる。

4. 結言

本稿では、調理動作の識別をするにあたり、切る動作に焦点を当て、画像識別による食材の種類と加工方法の同時識別の手法を提案した。画像識別により 1 種類のクラス分類をする一般的な CNN に対し、全結合層のみ種類ごとに独立させた構造にすることで、同時識別の実現を図った。実験により、同じ構造の 1 種類の場合における識別精度には劣るが、2 種類の同時識別において 87.9% と高い精度の識別結果を得ることができた。

次の段階として、食材の色や見た目が類似しているものや、加工後の形状が似ているが切り方の異なるもののデータを追加し、識別可能であるか検討する。また、切る動作以外においても画像による動作等の同時識別が可能であるか検討する。

謝辞 本研究の一部は NEDO の補助を受けたものである。

参考資料

- [1] Karinne Ramirez-Amaro, Emmanuel Dean-Leon, and Gordon Cheng, "Robust Semantic Representations for Inferring Human Co-manipulation Activities even with Different Demonstration Style", 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), pp. 1141-1146, 2015.
- [2] 蒯 承穎, 志土地 由香, 高橋 友和, 井手 一郎, 村瀬 洋, "料理映像における調理動作の解析", デジタルコンテンツシンポジウム講演予稿集, 4 巻, 2008.
- [3] Yuko Yamakata, Koh Kakusho, and Michihiko Minoh, "Object Recognition based on Object's Identity for Cooking Recognition Task", ISM 2010, 12th IEEE International Symposium on Multimedia, Taichung, Taiwan, 2010.