

AR DeepCalorieCam: AR 表示型食事カロリー量推定システム

丹野 良介[†] 會下 拓実[†] 柳井 啓司[†]

[†] 電気通信大学 大学院情報理工学研究科 情報学専攻

E-mail: [†]{tanno-r, ege-t, yanai}@mm.inf.uec.ac.jp

あらまし 近年、健康意識の高まりから、毎日の食事を記録するための様々な食品写真認識アプリケーションが提案されている。一部のアプリケーションでは、食品カテゴリの推定に画像ベースの分類を使用しているが、そのほとんどは物体検出を活用していない。これは、料理写真に複数の料理が含まれている場合に、これらの料理を個別に検出するためには、人の手によるアノテーション作業が必要であり、コストがかかることに起因する。しかし、現実には、食事をするシーンにおいて複数の料理を扱うことが一般的であるため、複数品目を対象にする場合は物体検出を用いるのが妥当である。本研究では、物体検出手法である YOLOv2 と Multi-task CNN を組合せた、複数品目食事画像の同時カロリー量推定システムを作成した。また一方で、スマートフォンの技術的進歩により気軽に拡張現実 (AR) を試すことが可能となっている。そこで、本研究では、深層学習によるカロリー量推定と AR 技術を組み合わせた AR 表示型食事カロリー量推定システム”AR DeepCalorieCam”を提案する。

キーワード カロリー量推定, 深層学習, アプリケーション, AR, 物体検出

1. はじめに

近年、健康意識の高まりにより、毎日の食事記録をつけるために様々な食事写真認識アプリケーションが提案されている。一部のアプリケーションでは、食事カテゴリの推定に画像ベースの分類を使用しているが、そのほとんどは物体検出を活用していない。これは、料理の写真に複数の食品が含まれている場合に、これらの食品を個別に検出するためには人間による個々の食品カテゴリ推定のための援助が必要であることを意味する。しかし、現実には、複数の料理を食べる場面が頻繁にあるため、物体検出を利用することが現実的である。一方、画像認識の分野では、CNN を用いた多くの手法が様々な改良を達成しており、現在、様々な分野で応用されている。

本論文では、CNN を使用した最先端の物体検出アルゴリズムである YOLOv2 [?] を使用して、食事写真から食品の検出を行う。YOLOv2 は、CNN を用いた state-of-the-art な物体検出ネットワークであり、高速かつ高精度な検出が可能である。YOLOv2 による物体検出は各物体の矩形領域を出力するため、同じカテゴリの個々の物体を認識することが可能である。複数の食品の食事写真の場合、これは、画像全体の各料理領域を推定することが可能であることを意味する。したがって、単一レベルの食事画像用に動作するシステムと、検出された各物体の矩形領域からより詳細に食品を認識することができる。

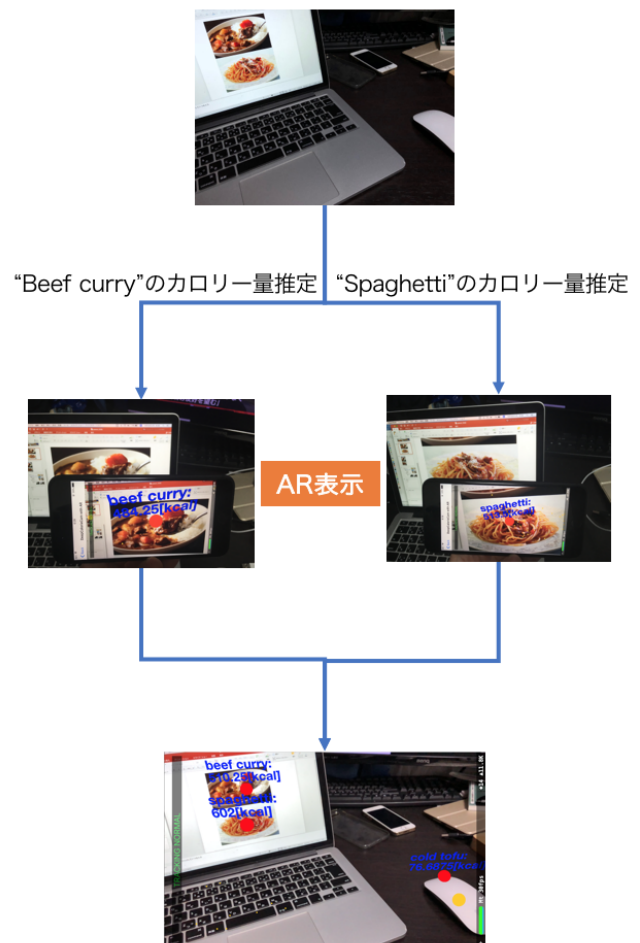


図1 AR 表示型食事カロリー量推定システムの処理フロー

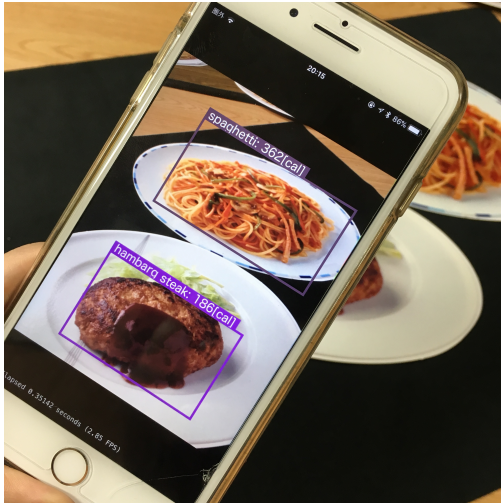


図 2 DeepCalorieCam



図 3 AR DeepCalorieCam

本論文では、iOS 上で実行される 2 つの自動カロリー量推定アプリケーション *DeepCalorieCam* と *AR DeepCalorieCam* を提案する。[1] は、画像ベースの食事写真の食品カテゴリとカロリー量の同時推定によるカロリー量推定を提案した。しかし、これは料理に利用される皿が 1 枚しか入っていない食事写真に限られているため、この食品検出器と組み合わせることで、複数の皿の食事写真から食べ物のカロリー量を推定することが可能となる。図 2 に iPhone 7 Plus で動作している *it DeepCalorieCam* の使用例を示す。また、*AR DeepCalorieCam* は、図 3 のように料理を検出するのではなく、カロリー推定と拡張現実感 (AR) の組合せから構成されるアプリである。

2. 提案システム

2.1 DeepCalorieCam

我々は、食事写真からカロリー量を推定するシステムを [?] で提案されている手法を用いて iOS アプリとして実装を行った。[?] では、画像ベースによる食事写真から食品カテゴリとカロリー量を同時に推定する手法を提案している。オンラインレシピ調理 Web サイトからカロリー量付きレシピデータを収集し、皿 1 枚のみを含む食事写真から直接、カロリー量を出力する Multi-task Convolutional Neural Network (Multi-task CNN) を学習している。食事カロリー量の推定に使用するネットワークは VGG16 [3] に基づいている。fc6 層は両方のタスクで共有され、fc7 層は各タスクに分岐している。各タスクは fc7 層と出力層を独立して持つ構成となっており、食事カロリー量推定タスクは、4096 次元の fc7 層と、食事カロリー量を出力す

る出力層を有する。食品分類タスクは、4096 次元の fc7 層と、各カテゴリに対応する単位で構成される出力層とを有する。

一方で、モバイルの実装を考慮すると、VGG16 は次の理由で不適切であると考えられる。(iPhone 8 Plus の推論時間、ImageNet Top-1 精度、円のサイズで表された学習済みモデルの容量の関係を示す Figure5 も参照)。

(1) モデルの重み (553 MB) のサイズは、モバイルの実装には大きすぎる

(2) 推論処理に時間を要する

特に、モバイルの実装では、デバイスのメモリ容量と処理時間は、深層学習を実装する上で重要な要素である。また、[?] の研究から、カロリー量推定の結果は分類精度に依存する傾向があることが示唆されている。現時点では、学習済みのモデルが小さく、より高速な推論を行うことができ、高い分類精度を誇る Inception-v3 [4] のネットワーク構造を Multi-task CNN の基本構造として使用することに決めた。

図 4 は食品領域検出とカロリー量推定システムの流れを示している。我々はこの食事カロリー量推定ネットワークと YOLOv2 [?] を組み合わせて、複数の料理から構成される食事写真から複数品目同時にカロリー量推定が行えるようにした。まず、複数の皿の食事写真から YOLOv2 による食器皿のバウンディングボックスを抽出し、1 つの皿のみを含む各バウンディングボックスに対応する食品領域画像を得る。次に、これらの切り抜かれた画像を 1 つずつ食事カロリー量推定ネットワークに与える。最後に、食事カロリー量の総量は、すべての切り抜かれた食品領域画像から推定された食事カロリー量から算出さ

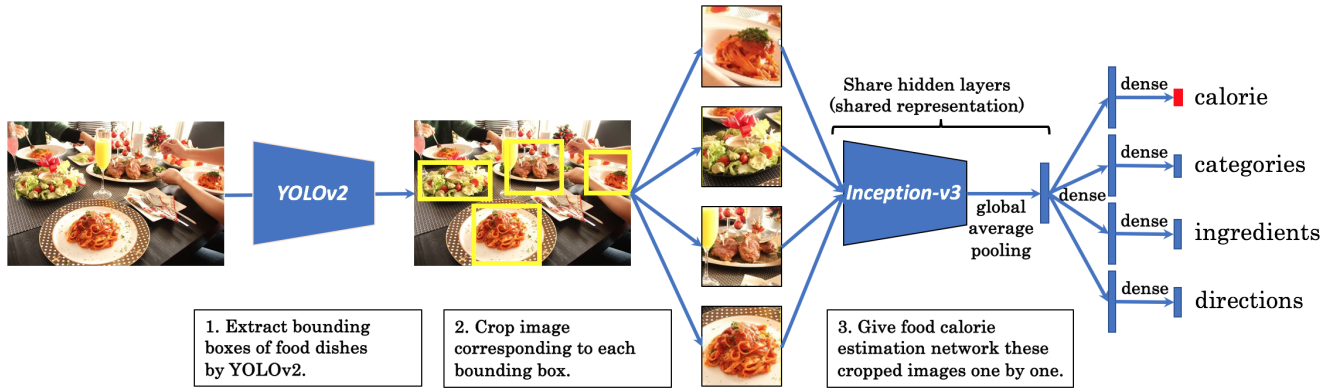


図 4 食事領域検出とカロリー量推定の処理フロー

れる。

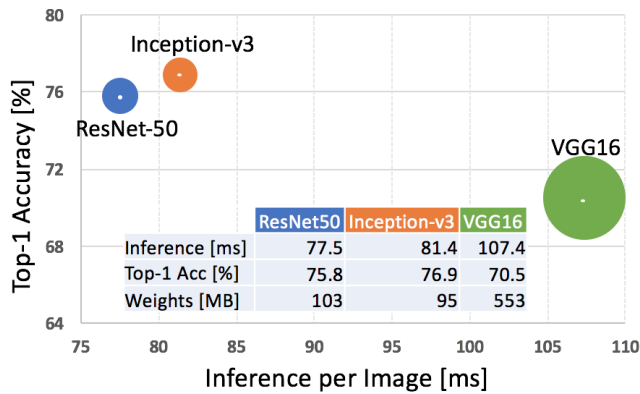


図 5 iPhone 8 Plus の推論時間, ImageNet Top-1 精度, 円のサイズで表された学習済みモデルの容量

2.2 AR DeepCalorieCam

本設では, AR 技術を用いた新しい食品認識手法を提案する. AR it DeepCalorieCam は YOLOv2 による食品領域の検出を行う部分を省略しているので, 同時に 1 つの食品しか認識することができない. しかし, 認識結果は 3D AR-view 空間に残っているため, スマートフォンを動かすことで連続した複数の食事認識及びカロリー量推定を実現した. 図 1 として AR 表示型食事カロリー量推定システムの処理フローを示す.

3. おわりに

本論文では, iOS 上で実行される 2 つの自動カロリー量推定アプリケーション *DeepCalorieCam* と *AR DeepCalorieCam* を提案した. [1] は, 画像ベースの食事写真の食品カテゴリーとカロリー量の同時推定によるカロリー量推定を提案した. しかし, これは料理に利用される皿が 1 枚しか入っていない食事写真に限られているため, この食品検出器と組み合わせるこ

とで, 複数の皿の食事写真から食べ物のカロリー量を推定することが可能となる. 図 2 に iPhone 7 Plus で動作している *it DeepCalorieCam* の使用例を示す. また, *AR DeepCalorieCam* は, 図 3 のように料理を検出するのではなく, カロリー推定と拡張現実感 (AR) の組合せから構成されるアプリである. 食事カロリー量の推定は, 食事写真を入力して食事カロリーの値を出力する回帰問題として扱われる.

一方, [2] は, 食品の 2D サイズから二次曲線 fitting によるカロリー量推定を提案している. この手法では, 予め登録された参照オブジェクトを用いて上から食事写真を取るだけで自動的に食事カロリー量の推定を行っている. 各食品カテゴリーの二次曲線は, 実際の食事カロリーでアノテーションされた学習データに基づいて学習されている. カロリー値は, 以下の式によって近似的に計算される.

$$Calorie = a_i * F_r^2 + b_i * F_r + c_i \quad (1)$$

ここで, a_i, b_i, c_i は, カロリー量のアノテーションがされた食事画像データセットから事前に学習された i 番目の食品カテゴリー毎に決まった数値からなる. また, a_i, b_i, c_i を最小 2 乗法で多項式近似法を用いて計算を行い, さらに, 実際の食品の大きさ, F_r は, 以下の式によって得ることができる:

$$F_r = S_r * \frac{F_p}{S_p} \quad (2)$$

ここで F_p は対象食品の領域のピクセル数を表し, S_p は参照オブジェクトの領域のピクセル数を表し, S_r はあらかじめ登録が予定されている基準物体の上面図の大きさを表している.

今後の課題としては, 回帰予測と実際の環境における二次曲線 fitting を用いて, 両者のカロリー量推定値を比較する予定である.

文 献

- [1] T. Ege and K. Yanai. Simultaneous Estimation of Food Categories and Calories with Multi-task CNN. In *Proc. of ACPR International Conference on Machine Vision Applications (MVA)*, 2017.
- [2] K. Okamoto and K. Yanai. An Automatic Calorie Estimation System of Food Images on a Smartphone. In *Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management (MADiMa)*, 2016.
- [3] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Proc. of International Conference on Learning Representations*, 2015.
- [4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. In *Proc. of arXiv:1512.00567*, 2015.