米飯を基準とした CNNによる食事画像からのカロリー量推定

會下 拓実[1,a)]　Jaehyeong Cho[1,b)]　松平 礼史[1,c)]　柳井 啓司[1,d)]

概要

近年, 様々な食事管理アプリケーションが開発され, 食事記録をすることが容易になっている. しかしこれらのアプリケーションでのカロリー量推定は, ユーザ入力が必要であったり, 栄養士を雇ったりと, 人手のかかるものとなっている. 一方, 画像認識分野では Convolutional Neural Network(CNN) を用いた手法が主なタスクの最高精度を独占している. この CNN による食事画像の認識に関する研究も盛んに行われているが, 高精度の食事画像からのカロリー量推定は実現されておらず, 現状では困難な問題である. そこで本研究では, 面積を考慮した食事画像からのカロリー量推定を行う. そのために, まず CNN を用いた米飯画像からの実寸推定を行う. 米飯粒は大きさが一定であるため, 複数の米飯粒が密集した米飯の画像から実寸を直接推定する CNN を構築する. そして料理領域分割と実寸推定を組み合わせることで, 面積を考慮した食事画像からのカロリー量推定を実現する. 実験では, 撮影した米飯画像に実寸情報を付与することで構築したデータセットを用いる. 実寸推定の実験を行った結果, 224 ピクセルあたりの実寸を推定したときの絶対誤差と相対誤差がそれぞれ 0.145cm と 5.548%となり, また, 推定値と正解値の相関係数が 0.946 となり, 高い相関が得られた.

1. はじめに

料理のカロリー量は料理カテゴリおよび量に強く依存すると考えられ, 料理カテゴリや量を食事画像から自動推定することが可能となれば, 食事管理の面で有用である. 食事画像からの料理カテゴリ分類においては既に CNN を用いた手法が高精度の分類を達成しており, 最近では画像認識により食事画像から料理名の候補を自動で提案するアプリケーションも存在する. しかしカロリー量計算に不可欠な料理の量においては, 料理カテゴリ毎に基準量を設けることで料理の量を考慮しないものも多い. また, 料理の量を考慮する場合であっても, ユーザ入力や基準物体が必要であるなど, 料理の量の取得は非常に困難である. 現状では, 食事画像からのカロリー量推定は未解決の問題となっている.

そこで本研究では面積を考慮した食事画像からのカロリー量推定を行う. そのために, まず CNN を用いた米飯画像からの実寸推定を行う. 料理の量を考慮した食事画像からのカロリー量推定の既存研究では, 対象の料理と同時に撮影された, 大きさが既知のカードなどの基準物体の領域と, 料理の領域を比較することで, 料理の面積に基づいてカロリー量を推定していた. これに対して本研究では, 大きさが一定である米飯粒が密集した米飯の画像から実寸を直接推定する. 米飯画像のパッチ画像を入力として, そのパッチ画像の一辺の長さの実寸を出力する CNN を構築する. そして料理検出と料理領域分割を行い, 実寸推定と組み合わせることで面積を考慮した食事画像からのカロリー量推定を行う. ただし本手法では, 食事画像をテーブル面に垂直に真上から撮影することを仮定する.

カードなどの明示的な基準物体の場合は，基準物体と同時に撮影する必要があり, そのような写真は意図的に撮影しない限り入手が困難であるが，米飯は，ご飯が含まれる食事の画像には多くの場合，写っているので，明示的に基準物体を含めた撮影をする必要がなく，ウェブ上にある食事写真や, 過去に撮影した食事写真もカロリー量推定の対象とすることが可能である.

まとめると, 本研究では面積を考慮した食事画像からのカロリー量推定を行う. そのために, まず米飯画像から実寸を直接推定する CNN を構築し, さらにその実寸推定と料理領域分割を組み合わせることで, 面積を考慮した食事画像からのカロリー量推定を行う.

2. 関連研究

食事画像からのカロリー量推定にはいくつかのアプローチが存在するが, 主要なアプローチは, 推定された料理カテゴリと料理の面積や体積の情報から, 事前に登録された料理カテゴリごとの単位面積当たりもしくは単位体積当たりのカロリー量の値を利用してカロリー量を推定する手法である.

Chen ら [1] は料理カテゴリを推定後, Kinect のような深度カメラにより料理の体積を推定し, 最終的にカロリー量

[1] 電気通信大学大学院情報理工学研究科
[a)] ege-t@mm.inf.uec.ac.jp
[b)] cho@mm.inf.uec.ac.jp
[c)] matsuda-r@mm.inf.uec.ac.jp
[d)] yanai@cs.uec.ac.jp

を推定している．深度カメラによる料理の体積の推定は正確であるが特殊なデバイスであるため，一般の人が普段使用することは難しいと考えられる．

Myers ら [2] が提案した Im2Calories は，CNN を用いて，食事/非食事の認識，複数品目の認識，深度推定，領域分割などの複数のタスクを行い，カロリー量を推定している．しかしカロリー量情報付きのデータセットの不足により，評価が十分に行われていないという問題がある．

岡元ら [3] は大きさが既知の基準物体と一緒に料理を撮影することで料理の体積を推定し，高精度のカロリー量推定を実現した．まず，基準物体と料理を一緒に撮影し，基準物体と料理のそれぞれの領域を抽出する．そして基準物体と料理の領域を比較して算出した料理の大きさからカロリー量を計算する．料理の領域の抽出では，まずエッジにより背景から皿領域を検出し，その皿領域に対して k-means により色情報に基づく領域分割を行い，最終的に GrabCut [4] により皿領域から料理領域を推定する．実験には基準物体と料理が一緒に写った画像が必要であるが，この手法は高精度の料理領域の推定を実現し，カロリー量推定では相対誤差 21%という精度を達成した．これに対して本研究では，CNN を用いて米飯画像からの実寸推定を行う．

3. 手法

本手法では，料理の面積を考慮したカロリー量推定を行うために，料理検出と料理領域分割，そして米飯画像からの実寸推定を行う．提案手法では以下の流れで面積を考慮したカロリー量の推定を行う．

(1) 対象の料理と米飯を同時に撮影
(2) CNN を用いて各料理を検出
(3) 各検出領域から CNN を用いて料理領域を抽出
(4) 米飯画像から実寸を推定
(5) 推定された実寸情報から各料理領域の実面積を推定
(6) 推定された実面積に基づきカロリー量を推定

提案手法では，食事画像をテーブル面に垂直に料理の真上から撮影することを仮定する．以下に各処理の詳細を記述する．

3.1 料理検出

本手法では各料理を検出するために Redmon らが提案した YOLOv2 [5] を使用する．YOLOv2 は以前に提案された CNN に基づく YOLO [6] を改善することで，高速かつ高精度な物体検出を可能としている．本手法では，YOLOv2 の学習には，大規模食事画像画像データセットである UECFood-100 [7] に含まれる画像に付与し直した料理バウンディングボックスを使用する．画像 5000 枚を用いて YOLOv2 の学習を行い，500 枚での評価を行った結果，AP (Average Precision) が 0.8 となり，高精度な料理検出を可能とした．図 1 に，付与し直した料理バウンディングボックスを学習した YOLOv2 での料理検出の結果と，UECFood-100 を学習したクラス分類モデルにより各検出領域から推定された料理カテゴリを示す．

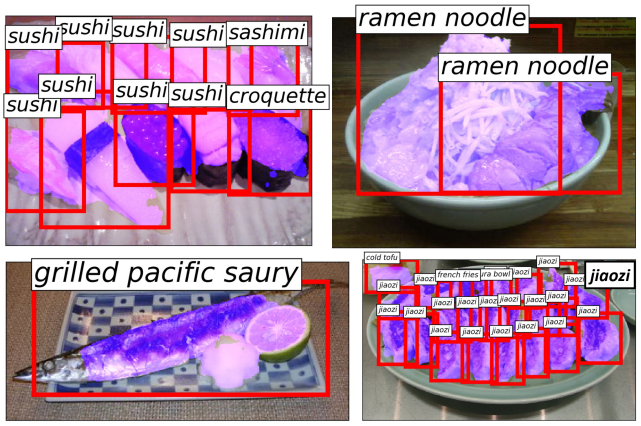

図 1 料理検出と料理領域分割の結果．赤枠が推定されたバウンディングボックス，各赤枠上のタグが推定された料理カテゴリ，赤い領域が推定された料理領域である．

3.2 料理領域分割

本手法では各料理の領域を推定するために，Ronneberger らが提案した U-Net [8] を使用する．U-Net は入力層付近の解像度の高い特徴マップを出力層付近に用いることで，高精度な領域分割を可能としている．本手法では，U-Net の学習には，UECFood-100 [7] に含まれる画像に新たに付与した料理セグメンテーションマスクを使用する．画像 5000 枚を用いて U-Net の学習を行い，500 枚での評価を行った結果，IoU (Intersecting over Union) が 0.8 となり，高精度な料理領域分割を実現した．図 1 に，新たに付与した料理セグメンテーションマスクを学習した U-Net での料理領域分割の結果を示す．

3.3 米飯画像からの実寸推定

本手法では面積を考慮した食事画像からのカロリー量推定を行うために，CNN による米飯画像からの実寸推定を行う．大きさが一定である米飯粒が密集する米飯の画像から実寸を直接推定する CNN を構築する．図 2 に本手法の実寸推定モデルの構造を示す．

本研究で用いる CNN は VGG16 [9] に基づく．本手法の実寸推定モデルは図 2 のように，実寸を出力する単一のユニットで構成される出力層を有する．入力は米飯画像から得られるパッチ画像であり，出力は入力されたパッチ画像の一辺の長さの実寸である．本実験では入力パッチ画像のサイズを 224×224 とし，出力は 224 ピクセルあたりの実寸となる．

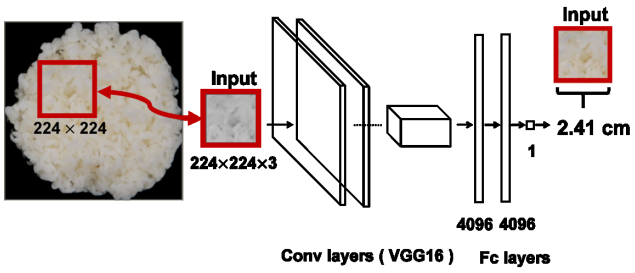

図 2 米飯画像から実寸を推定する CNN の構造．

3.4 カロリー量推定

本手法では, 料理検出と料理領域分割により得られた料理領域情報と, 米飯画像から推定された実寸情報から各料理の実面積を求め, そこから岡元ら [3] の手法に従い, 各料理のカロリー量を推定する. 岡元らは事前に複数サイズの料理のカロリー量より学習した回帰曲線に基づき, 対象食品 20 種類に関して実面積情報からカロリー量を推定した. 料理カテゴリ毎に学習された回帰曲線を用いることで, 牛丼やごはんといった深さのある皿に盛られることが多い料理に関しても 2 次曲線のようなフィッティングを行うことが可能である. 本手法では, 岡元ら [3] が収集した画像から求めた回帰曲線に基づき, 実面積情報からカロリー量の推定を行う.

4. データセット

本研究では米飯画像からの実寸推定を行うために, 実寸情報付き米飯画像データセットを新たに作成した. まず米飯画像の撮影では, 2 種類のカメラ (COOLPIX AW120 と iPhone8 Plus) を使用し, 各水量 (米 150g に対して水 180ml, 200ml, 220ml) を用いて炊飯したそれぞれの米飯の撮影を行った. カメラの種類と炊飯時の水量の組み合わせごとに 60 枚撮影したため, 全体で 360 枚の画像が収集された. また, 1 枚撮影する毎に被写体の米飯とカメラとの間の距離を変更し, 5 枚撮影する毎に米飯の盛り付けを変更することで, 多様な米飯画像を収集した. 次に撮影した各米飯画像に 1 ピクセルあたりの実寸情報を付与し, さらに背景情報を除去するために米飯領域マスクの作成を行った.

5. 実験

本実験では, 米飯画像からの実寸推定と面積を考慮した食事画像からのカロリー量推定を行う. 米飯画像からの実寸推定での学習と評価には, 本研究で作成された実寸情報付き米飯画像データセットを用いる. 面積を考慮したカロリー量推定では, 料理領域分割と実寸推定を組み合わせることで, 実面積を推定し, そこからカロリー量を推定する. テスト画像として, UECFood-100 [7] の複数品料理画像のうち, 画像に米飯が含まれるものを用いる.

5.1 米飯画像からの実寸推定

本実験では米飯画像からの実寸推定を行う. カメラの種類と炊飯時の水量の組み合わせに基づき米飯画像データセットを 6 つに分割し, 5 つを学習に使用し, 残りの 1 つを評価に使用する. したがって学習画像と評価画像はそれぞれ 300 枚と 60 枚となる.

学習時には, 米飯画像のランダムな位置から切り抜かれた 1 枚の 224×224 のパッチ画像から推定された実寸に対して学習が行われる. また, 前処理として米飯画像の拡大縮小と左右反転, 回転を行う. 拡大縮小では, ある米飯画像を n 倍に拡大縮小した場合, その米飯画像に付与された実寸情報を $\frac{1}{n}$ 倍にする. 評価時には, 1 枚の米飯画像から 5×5 のグリッドサンプリングによって切り抜かれた 25 枚のパッチ画像それぞれから推定された実寸の平均値が最終的な出力となる. さらに学習時と評価時の両方において, 付与された米飯領域マスクに基づき, 背景領域が 1%以上を占めるパッチ画像は背景画像として除去される.

本実験において米飯画像からの実寸推定に用いる CNN の構造は VGG16 [9] に基づく. 出力層以外の層では, ImageNet の 1000 種類分類タスクにより学習済みの重みを学習時の初期値として使用する. 最適化手法として SGD を使用し, momentum 値を 0.9 とする. バッチサイズを 16 として学習率 10^{-5} において約 3,000 回反復する. また, 損失関数として 2 乗和誤差を使用する.

評価指標として絶対誤差, 相対誤差, 推定値と正解値の相関係数, 相対誤差 5%, 10%, 20%以内の推定値の割合を用いる. 絶対誤差は推定値と正解値の差の絶対値であり, 相対誤差は正解値に対する絶対誤差の割合である. なお, 評価には 224 ピクセルあたりの実寸における推定値と正解値を用いる.

表 1 に実寸推定の結果を示し, 図 3 に推定値と正解値の相関を示す. 224 ピクセルあたりの実寸を推定したときの平均絶対誤差と平均相対誤差がそれぞれ 0.145cm と 5.548%となり, また, 推定値と正解値の平均相関係数が 0.946 となり, 高い相関が得られた.

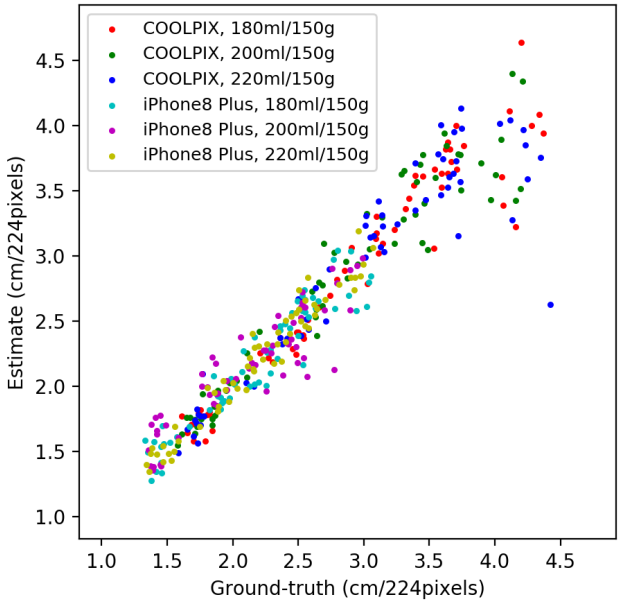

図 3 米飯画像からの実寸推定における推定値と正解値の相関関係.

学習データと評価データの全ての組み合わせにおいて, 相対誤差が 10%より小さく, 相関係数が 0.9 より大きくなり, また, ほとんどの推定値が相対誤差 20%以内 (%) に含まれている. 本手法では, 米飯画像から実寸を直接推定する CNN を学習することで, 米飯粒ひと粒ひと粒の大きさや向きを考慮せずに推定を可能にしたと考えられる.

表 1　米飯画像からの実寸推定 (224 ピクセルあたりの実寸における評価)

評価データ	絶対誤差 (cm)	相対誤差 (%)	相関係数	相対誤差 5%以内 (%)	相対誤差 10%以内 (%)	相対誤差 20%以内 (%)
COOLPIX, 180ml/150g	0.152	4.822	0.963	61.667	88.333	98.333
COOLPIX, 200ml/150g	0.169	5.513	0.959	55.000	85.000	100.000
COOLPIX, 220ml/150g	0.194	5.906	0.920	55.000	86.667	96.667
iPhone8 Plus, 180ml/150g	0.123	5.706	0.949	51.667	85.000	100.000
iPhone8 Plus, 200ml/150g	0.145	7.305	0.910	56.667	66.667	91.667
iPnone8 Plus, 220ml/150g	0.086	4.037	0.976	71.667	95.000	100.000
Average	0.145	5.548	0.946	58.611	84.444	97.778

5.2　面積を考慮したカロリー量推定

本実験では CNN による料理領域分割と米飯画像からの実寸推定を組み合わせることで, 領域分割に基づく面積を考慮した食事画像からのカロリー量推定を行う. YOLOv2 [5] での料理検出により得られた各料理領域に対して, クラス分類と U-Net [8] による領域分割を行うことで各料理のカテゴリと領域を推定する. さらに抽出した米飯画像から実寸を推定することで得られた 1 ピクセル当たりの実面積を基準として, 料理の実面積を算出し, そこから岡元ら [3] の手法に従い, カロリー量を推定する. テスト画像として, UECFood-100 [7] の複数品料理画像のうち, 画像に米飯が含まれるものを用いる. 図 4 に実面積推定とカロリー量推定の結果を示す.

図 4　米飯を含む複数品料理画像からの面積を考慮したカロリー量推定の結果. 各赤枠上のタグが推定された実面積とカロリー量である.

6.　おわりに

本研究では料理面積を考慮した食事画像からのカロリー量推定を行うために, CNN による米飯画像からの実寸推定を行った. 米飯粒は大きさが一定であるため, 複数の米飯粒が密集した米飯の画像から実寸を直接推定する CNN を構築した. 実験では 224 ピクセルあたりの実寸を推定したときの平均絶対誤差と平均相対誤差がそれぞれ 0.145cm と 5.548%となり, また, 推定値と正解値の平均相関係数が 0.946 となり, 高い相関が得られた. そして料理検出と料理領域分割を行い, 米飯画像からの実寸推定と組み合わせることで, 面積を考慮した食事画像からのカロリー量推定を行った.

今後の課題として, 面積を考慮したカロリー量推定の評価を行うことがある. 評価を行うために, 米飯が含まれるカロリー量情報付き複数品料理画像データセットを作成する予定である. また, 米飯がない状況においても実面積推定が可能なシステムを構築するために, 領域分割と基準物体を用いる手法 [3], [10] や, iPhone の深度対応カメラから得られる深度情報と組み合わせることも検討している.

参考文献

[1] M. Chen, Y. Yang, C. Ho, S. Wang, E. Liu, E. Chang, C. Yeh, and M. Ouhyoung. Automatic chinese food identification and quantity estimation. In *Proc. of SIGGRAPH Asia Technical Briefs*, pp. 1–4, 2012.

[2] A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and P. K. Murphy. Im2calories: towards an automated mobile vision food diary. In *Proc. of IEEE International Conference on Computer Vision*, pp. 1233–1241, 2015.

[3] 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*, 2016.

[4] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, Vol. 23, No. 3, pp. 309–314, 2004.

[5] J. Redmon and A. Farhadi. YOLO9000: Better, faster, stronger. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2017.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You Only Look Once: Unified, real-time object detection. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2016.

[7] Y. Matsuda, H. Hajime, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 25–30, 2012.

[8] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Springer*, pp. 234–241, 2015.

[9] K. Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. In *arXiv preprint arXiv:1409.1556*, 2014.

[10] W. Shimoda and K. Yanai. CNN-based food image segmentation without pixel-wise annotation. In *Proc. of IAPR International Conference on Image Analysis and Processing*, 2015.