

全体カロリー量のみがアノテーションされた 複数品食事画像の個別カロリー量推定

岡本 開夢^{1,a)} 足立 賢人^{1,b)} 柳井 啓司^{1,c)}

概要

深層学習の登場によって画像認識の性能が向上し、最近では食事画像からの食事カロリー量推定の研究が広く行われている。食事画像は Web 上に大量に存在するため容易に大量の画像を集めることはできるが、カロリー量推定に必要な個々の食事カテゴリのカロリー量のアノテーションは Web 上にはほとんど存在せず、データセットの構築が研究のボトルネックとなっている。本研究では、我々が Web 上で発見した、複数の食事が含まれているがその合計の食事カロリー量のみがアノテーションされているデータを利用して、食事カロリー量推定の学習を実現する方法について検討する。具体的には、全体カロリー量と領域推定手法を組み合わせ、各食事カテゴリのカロリー量を推定するモデルを提案し、他のモデルと比較実験により最も柔軟に各食事カテゴリのカロリー量を推定可能であることを示す。

1. はじめに

現在、深層学習の発達により画像認識の精度が飛躍的に向上し、画像生成や領域分割といったタスクにおいても優れた成果を残している。その成果は、医療や自動運転といった実生活にも用いられ我々の生活を豊かなものにすることに貢献している。また、食生活に関しても、画像認識を応用した自動食事記録アプリなどが登場し、さらに高度な応用として、深層学習による食事画像からのカロリー量推定が行われている。これは、食事の見た目とカロリー量の相関性を用いており、画像から食事領域を推定し、それぞれの領域のカロリー量を推定することで、食事全体の総カロリー量を推定する。2次元、3次元ベース両方の手法で行われており、2次元ベースの手法によるカロリー量推定には食事領域分割が不可欠であるといえる。

多くのカロリー量推定には個々の食事カテゴリにおけるカロリー量が必要となるが、実生活における栄養素の表記においては全体のみを扱っている場合が多い。そこで本研

究では、全体カロリー量付きのセグメンテーションマスク付き給食データセットを用いた各食事カテゴリのカロリー量推定手法の提案を行う。データセットとしては、我々が Web 上で発見した、複数の食事が含まれているがその合計の食事カロリー量のみがアノテーションされている食事画像データにさらに領域アノテーションを付与して利用する。本研究では、領域推定モデルにランチを追加することにより、領域推定と合計カロリー量を組み合わせることで各食事カテゴリのカロリー量の推定を実現する。他のモデルと比較実験を行い、提案するモデルが最も柔軟に各食事カテゴリのカロリー量を推定可能であることを示す。

2. 関連研究

カロリー量に関する研究として、CNN 以前では GrabCut [10] を用いた手法が行われており、岡元ら [6] はモバイル上でのカロリー量推定を実現した。これは、大きさが既知の基準物体とともに食品を上から撮影することでカロリー量を推定するシステムである。GrabCut [10] を用いて画像内の基準物体と食事の領域を抽出することで実面積を算出し、事前に作成したカロリーと面積の回帰式をもとにカロリー量の推定を行っている。しかし、この研究は撮影角度が上からのみと限定的なものであった。CNN 登場後のカロリー量推定に関する研究として、安藤らの DepthCalorieCam[1] がある。これは、モバイル上で U-Net[9] による食事領域を抽出したのち、iPhone に内蔵されている深度情報を用いることで体積を推定しカロリー量を推定している。この研究ではカロリー量を推定できる食事カテゴリは3種類となっており限定的である。一方で、複数品目の各食事カテゴリごとカロリー量を推定した研究として會下らの研究 [3] がある。これは、Faster R-CNN [7] によって複数品目を含んだ画像内の各食事を検出し、事前に学習した VGG [11] を用いて各食事カテゴリごとのカロリー量推定を行なっている。また、會下ら [4] は食事領域分割用データセットを用いたカロリー量推定も行なっている。これは、米飯の米粒が一定サイズという仮定のもと画像内の米飯領域の実寸を推定しており、参照物体なしで米飯画像を含んだ食事領域の面積推定を可能にしている。推

¹ 電気通信大学院 情報理工学研究科 情報学専攻

a) okamoto-ka@mm.inf.uec.ac.jp

b) adachi-k@mm.inf.uec.ac.jp

c) yanai@cs.uec.ac.jp

定された面積をもとに事前に作成しておいたカロリー量/面積の回帰式をもとにカロリー量を推定している。

3. 手法

現在、既存のカロリー量推定は推定される個々の食事カテゴリの基準となるカロリー量が既知であることを前提として行われている。例えば、“ピザ”のような単品目のみが含まれている画像を扱う場合、基準となるカロリー量は一つ用意すればいいが、“定食”のような“ごはん”、“味噌汁”や“焼き魚”などの複数品目を含む画像を扱う場合、食事カテゴリの数だけカロリー量の正解値を用意する必要がある。しかし、実生活における栄養素の表記においては食事全体のみを扱っている場合が多く、個々のカロリー量を知る機会は少ない。実際に、Web 上には全体カロリー量のみが記載された給食画像、コンビニの弁当画像、ファミリーレストランのメニュー画像などが存在している。そこで、本章では食事画像に含まれている個々の食事カテゴリのカロリー量の推定を行う。カロリー量と食事領域は密接に関係していることから、今回カロリー量推定において食事領域の推定結果を利用してカロリー量の推定を行う。なお、今回は、給食画像を利用するため、すべての画像がほぼ同一のスケールで撮影されていることを想定する。画像毎にスケールが異なり、実寸の推定が必要な場合は、例えば、スマートフォンのステレオカメラ [1] や大きさ既知の基準物体 [4] など方法を利用する必要がある。

3.1 データセット作成

今回、食事領域推定とカロリー量推定を同時に行うため全体のカロリー量が既知なセグメンテーションマスク付きの食事データセットが必要となる。しかし、ベンチマークとなるような全体カロリー量が既知なセグメンテーションマスク付きの食事データセットは存在しない。そこで、新たに全体カロリー量が既知なセグメンテーションマスク付きの食事データセットを作成する。まずはじめに、會下ら [3] は複数品目のカロリー量を推定するために、印西市の給食センターの web サイト *1 から収集した給食画像を使用している。このサイトは毎日配給されている給食の献立、画像と全体カロリーをアップロードしているため、全体カロリー量が既知な食事画像データセットとなっている。そこで、新たに全体カロリー量が既知なセグメンテーションマスク付きの食事データセットを作成する。食事カテゴリ数は、同時にアップロードされている献立を参照した 60 カテゴリとした。データセットの画像例は、図 1 のとおりである。このデータセットは 593 枚の給食画像を含んでおり、学習画像を 474 枚、評価画像を 119 枚として用いる。平均値が 640.5kcal、分散が 36.7 である。



図 1 印西市の給食画全体カロリー量付きのセグメンテーションマスク付き給食データセット (60 カテゴリ)。

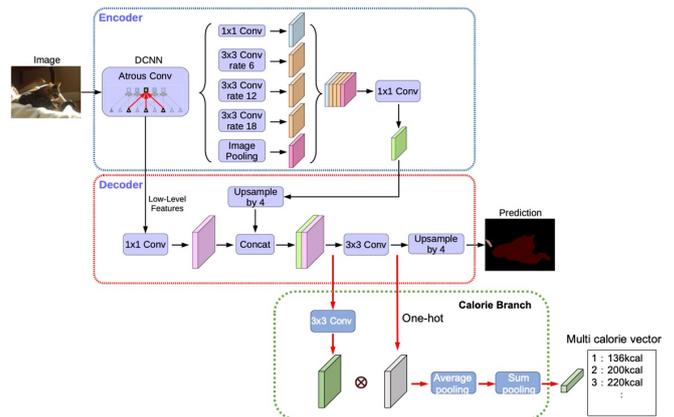


図 2 領域推定とカロリー量推定を組み合わせた提案モデルの概要図 ([2] から引用, 改変)。

3.2 手法概要

ここでは、領域推定と組み合わせることで食事画像から個別の食事カロリー量を推定を行う。今回、既存の領域推定モデルである Deeplab V3+ [2] に、新たなカロリー量推定ブランチを追加することを提案し各食事カテゴリのカロリー量を推定する。このブランチでは、画像特徴量から作成されたカロリー量特徴マップと Deeplab V3+ のセグメンテーション部分から算出されたセグメンテーション特徴マップを組み合わせることで食事カテゴリごとのカロリー量ベクトル v_c を推定する。今回カロリー量推定に用いる全体のモデルの概要図は図 2 の通りである。モデルは 2 段階に分けて学習され、はじめに Deeplab V3+ を用いて領域推定部分の学習を行った後、学習した領域推定部分のパラメータを固定しカロリー量推定部分の学習を行う。算出されたカロリー量は食事カテゴリごとにカロリー値を持っており、足し合わせた合計カロリーを用いて学習を行う。

3.3 Deeplab V3+ の学習

まず、はじめに領域推定と組み合わせるカロリー量推定を行うために、Deeplab V3+ モデルの学習を行う。モデルのバックボーンを ResNet-101 [5] として行う。作成した給食データを用いて学習を行った。評価には各クラスにおける Accuracy と mIoU を用いておりそれぞれ 0.61, 0.48 であった。カロリー量推定の際には、食事カテゴリ同士の関係と位置情報を保証し学習するため、学習した Deeplab

*1 http://inzai.ed.jp/kyusyoku/?page_id=32

V3+部分のパラメータを固定して用いる。

3.4 カロリー量推定ブランチ

カロリー量推定ブランチでは、カロリー量の特徴マップ $T_{calorie}$ とセグメンテーション特徴マップから生成されるチャンネル方向が one-hot なベクトルで構成されるテンソル $T_{one-hot}$ を用いてカロリー量ベクトル v_c を算出する。

$T_{calorie}$ は、Deeplab V3+のエンコーダーから算出された画像特徴量からセグメンテーション特徴マップと同様の方法で算出される。つまりは、画像特徴量を2つの $3 \times 3 \times \text{Conv}$ 層に通すことで畳み込み、 $1 \times 1 \times \text{Conv}$ 層によってチャンネル数を調整され、 $C \times W \times H$ のサイズを持つカロリー特徴マップとなる。ここでの C はデータセットのカテゴリ数に相当し、このカロリー量の特徴マップはセグメンテーション特徴マップと同サイズである。提案モデルでは、カロリー量特徴マップ $T_{calorie}$ が、ピクセルあたりの単位カロリー量をカテゴリ毎に表す特徴マップとなることを想定している。

セグメンテーション特徴マップは、チャンネル方向に softmax 関数を適用し最大値となるチャンネルを選択することによって、各ピクセルにおけるカテゴリを決定しマスク画像を生成している。そこで、セグメンテーション特徴マップをチャンネル方向に softmax 関数を適用し、最大値を持つチャンネルの値を1としてそれ以外のチャンネルの値を0とすることで、チャンネル方向に one-hot な値をもつテンソル $T_{one-hot}$ を作成する。これにより、画像内に同時に存在するカテゴリ間の関係を空間的に保持することができる。また、特徴マップのサイズを変更する処理を含まないため、カロリー特徴マップと同じサイズである。 $T_{calorie}$ と $T_{one-hot}$ のアダマール積をとり、Average pooling と Sum pooling を行うことで、各食事カテゴリのカロリー量ベクトル v_c を算出する。このブランチの $T_{one-hot}$ は、セグメンテーション特徴マップの精度に依存している。そのため、学習したセグメンテーション部分のパラメータを固定し $T_{one-hot}$ を算出することで、各食事カテゴリの位置関係を保証して用いる。

3.5 カロリー量推定ブランチの損失関数

カロリー量推定ブランチでは、個々の食事カテゴリのカロリー量をもつベクトル v_c が算出される。それぞれの食事カテゴリのカロリー量を足し合わせた合計カロリー量 v と、正解となる合計カロリー量を g を用いて、カロリー量推定ブランチの学習を行う。学習に用いる損失関数 L_{cal} は、重さ付きの絶対誤差 L_{ab} と相対誤差 L_{re} を用いる。

重み付き絶対誤差では、推定カロリー量が0に近い値にならないようにするために、推定値 v が正解値 g よりも小さい場合は重みを大きくした。次のように定義される。

$$L_{ab} = \lambda |v - g| \begin{cases} \lambda = 1.0 & (v - g \geq 0) \\ \lambda = 1.2 & (\text{otherwise}) \end{cases}$$

相対誤差 L_{re} は次のように定義される。

$$L_{re} = \frac{|v - g|}{g}$$

全体の損失関数は次のように定義される。

$$L_{cal} = \lambda_{ab} L_{ab} + \lambda_{re} L_{re}$$

λ_{ab} と λ_{re} はそれぞれの損失にかかる重みであり、実験では経験的に $\lambda_{ab} = 0.1$ と $\lambda_{re} = 0.01$ とし学習を行う。

4. 実験

4.1 各モデルの学習と評価

次に、合計カロリー量から各食事カテゴリのカロリー量の推定を行う。ベンチマークテストとして3つのモデルを用意し比較実験を行う。モデル A を食事画像から単に各食事カテゴリのカロリー量を推定する回帰モデルとし、VGG16 [11] を改良したモデルとする。モデル B をマルチラベルとモデル A を組み合わせたモデルとする。マルチラベルの学習を行った VGG16 の Conv 層のあとに Fc 層を用いた各食事カテゴリごとのカロリー量を推定するブランチを追加し、マルチラベル分類の結果とカロリー量推定ブランチの結果を掛け合わせることで各食事カテゴリのカロリー量を推定する。モデル C を提案したカロリー量推定ブランチを追加した Deeplab V3+モデルとする。合計カロリーはそれぞれのモデルで推定した各食事カテゴリごとのカロリー量を合計した値とする。また、會下らの手法も参考値として記述する。参考値とするのは、會下らの手法では、カテゴリ毎にカロリー量が付与された別の画像データセットを用いて学習し、給食データセット自体は評価用としてのみ使っているため正確な比較が出来ないためである。

学習と評価には、作成した給食データセットを用いる最適化手法として、Stochastic Gradient Descent(以下 SGD) [8] を用い、Momentum 値は 0.9、学習率は 10^{-4} とする。バッチサイズは 4 とし 200 エポック反復学習を行う。

4.2 各食事カテゴリのカロリー量推定結果

結果の指標として、推定された合計カロリー量の絶対誤差と各食事カテゴリのカロリー量の平均値を算出する。各食事カテゴリのカロリー量は推定されたカテゴリのみを対象とした。また、推定された各食事カテゴリの参照カロリー量は、厚生労働省の Web ページにて掲載されている食生活チェックブック*2と文部科学省の食品成分データベース*3を参考にした値である。食事のサイズはカロリー量と

*2 <https://www.mhlw.go.jp/bunya/kenkou/pdf/eiyousyokujij8.pdf>

*3 <https://fooddb.mext.go.jp>

表 1 合計カロリー量の平均絶対誤差 (kcal).

モデル	絶対誤差 (kcal)	相対誤差 (%)	20(%) ≤ 相対誤差	40(%) ≤ 相対誤差
モデル A (回帰)	45.0	7.3	99.2	100
モデル B (マルチラベル)	44.2	6.9	96.1	100
モデル C (領域推定)	74.8	14.6	80.7	96.6
参考モデル (會下ら)	136	21.4	53.0	85.1

表 2 各モデルによるカロリー量の平均推定値の抜粋結果.

食事カテゴリ	モデル A	モデル B	モデル C	参照カロリー量
牛乳	11.5	457.6	123.3	130
ご飯	12.1	50.9	209.4	250
まぜご飯	10.6	21.7	220.2	250
食パン	10.8	7.5	190.3	220
和風和え	11.9	33.1	60.2	50
グリーンサラダ	11.9	33.0	68.9	50
豚ロース焼き	12.2	2.5	74.4	240
鶏の照り焼き	11.6	2.0	35.8	300
味噌スープ	10.3	59.9	147.6	159
ミネストローネ	11.1	15.5	141.4	159
フルーツポンチ	11.1	5.2	68.1	50
みかん	9.6	0.5	45.5	50



図 3 給食データのカロリー量推定結果.

グラムを掲載している Web ページ^{*4}を参考にした. それぞれの結果に関しては表 1 と表 2 の通りである. さらに, 図 3 に 4 枚のテスト画像に対する個々の食品領域検出結果と推定カロリー量を示す.

まず, 合計カロリー量の平均絶対誤差ではモデル A, モデル B が良い精度となり, 提案手法が最も低い値となっている. これは給食画像の合計カロリーが 650kcal の前後であり推定しやすいデータセットであるためと考えられる. そのため, 画像間に含まれるカテゴリ間の情報を考慮せずとも単に近しいカロリー量を推定するだけで精度が良い結果となるため, 単純なモデルであっても高い精度で推定することが可能であった. 合計カロリー量にばらつきのあるデータセットを使用した場合には, より精度が違ったものになると考えられる.

一方で, 各食事カテゴリごとのカロリー量の値に関しては, 提案手法であるモデル C が最も良い精度となった. これは, 提案手法であるモデル C では各カテゴリ間の関係とその領域の大きさを考慮したため各食事カテゴリにばらつきを含んだカロリーを推定することができたためであると考えられる. モデル A では明示的なラベル間の関係を学習することができないため, 均一な値しか算出することができない結果となっている. モデル B では特定のカテゴリ, 特に牛乳が合計カロリー量の大部分を占める結果となった. これは各ラベル間の関係しか考慮していない他, マルチラベルだけでは少数の枚数しか含まないカテゴリの情報をカロリー量に反映することができていないためと考えられる. これらのことから, 各食事カテゴリごとのカロリー量推定には領域推定を用いることでより適切な値を推定することが分かった. 今回の実験では“豚ロース焼き”や“鶏の照り焼き”といった肉類に関する推定精度が低いため, 今後はデータセットを大規模にすることで, 推定精度の向上を目指す.

5. おわりに

本研究では, 領域推定モデルと合計カロリー量を組み合わせることで各カテゴリにおけるカロリー量の推定を可能にした. 今後の課題として, データセットの拡張や作成によってより汎化性能を高めたモデルの提案と実寸を考慮した手法の提案が挙げられる.

参考文献

- [1] Ando, Y., T., E., Cho, J. and Yanai, K.: DepthCalorieCam: A Mobile Application for Volume-Based Food Calorie Estimation using Depth Cameras, *MADiMA* (2019).
- [2] Chen, L., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, *ECCV* (2018).
- [3] Ege, T. and Yanai, K.: Estimating Food Calories for Multiple-dish Food Photos, *ACPR* (2017).
- [4] Ege, T. and Yanai, K.: Multi-task learning of dish detection and calorie estimation, *MADiMA* (2018).
- [5] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *CVPR* (2016).
- [6] Okamoto, K. and Yanai, K.: An Automatic Calorie Estimation System of Food Images on a Smartphone, *MADiMA* (2016).
- [7] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks, *PAMI*, Vol. 39, No. 6, pp. 1137–1149 (2017).
- [8] Robbins, H. and Monro, S.: A stochastic approximation method, *The annals of mathematical statistics*, pp. 400–407 (1951).
- [9] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation., *MICCAI*, p. 234–241 (2015).
- [10] Rother, C., Kolmogorov, V. and Blake, A.: “GrabCut”: Interactive Foreground Extraction Using Iterated Graph Cuts, *ACM Trans. Graph.*, Vol. 23, No. 3, pp. 309–314 (2004).
- [11] Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, *ICLR* (2015).

*4 <https://www.eiyoukeisan.com>