

## 深層学習を用いたスーパーでの様々な商品の重量推定 Weight Estimation of Various Products in Supermarkets Using Deep Neural Networks

菅 シャフィウルハサン<sup>†</sup>  
Shafiuruhasan Kan

青木 輝勝<sup>†</sup>  
Terumasa Aoki

### 1. 研究背景と目的

ロボットが行動決定をする際に自身の周りにある物体の重量を推定する必要がある。物体の重量を推定するというタスクは、ロボットにセンサーを取り付けたうえで、ロボットが対象の物体を持ち上げることで推定するという方法も可能である。しかし、画像のみから目の前にある様々な物体の重量を瞬時に推定できれば、作業効率が大幅に向上する。

画像を用いた物体の重量推定の研究は多く存在するが、そのほとんどは、例えば、豚やレタス、卵などというように特定の種類の物体に対する重量推定の研究である[1][2]。それらの研究においては、それぞれの物体が持つ固有の特徴を利用した重量推定の手法が提案されており、その手法を他の物体の重量推定に適応させることは難しい。

筆者らの知る限りでは、様々な物体の重量推定を試みた研究は2つ存在しか存在しない[3][6]。また、それらの研究の手法の精度を高くするために、モデルの一部を変更することを提案した研究がそれぞれ1つずつ存在する[4][7]。

しかし、これらの研究は画像の他に体積や密度を入力することが前提となっている。実際の現場ではこれらの情報が確実に入手できるとは限らないため、単一画像のみからの重量推定が必要であると考えられる。本研究では、単一画像のみから CNN などを利用して重量推定を行う手法を提案する。

### 2. 関連研究

Standley らはオンラインショップで販売されている様々な種類の物体に対して重量推定を行うモデルである image2mass を考案した[3]。image2mass では、オンラインショップにて API を使用して多くの商品の中から不備のない物を厳選して集めたデータセットが使われている。このデータセットは商品の画像 1 枚に加え、そのサイズ(width, height, length)、その重量がセットになったものである。image2mass は大きく分けて、Geometry Module、Volume Tower、Density Tower と名付けられた 3 つの部分で構成されている。Geometry Module と Volume Tower では物体の体積を推定し、Density Tower では物体の密度を推定し、それぞれから出力されたスカラー値を乗算することで重量を推定している。なお、Geometry Module のみ別のデータセットを使って事前学習されている。図 1 は image2mass のモデル図である。

また、Andrade らは Pix2Vox++ というモデルを使用して image2mass モデルの精度向上を実現した[4]。Pix2Vox++ は画像から画像内の物体の形状を推定することができる[5]。

これを利用して、image2mass の体積推定の精度向上を可能にした。

様々な種類の物体に対する重量推定の研究の他の例として、Dalai らの研究がある[6][7]。彼らの提案した手法は、画像から物体の種類のカテゴリと体積の推定を行い、事前に用意してあった物体の種類ごとの密度のデータベースから密度を検索し、得られた密度と体積を乗算するというものである。

筆者らの知る限りでは、様々な種類の物体に対する重量推定の研究は上記の報告のみである。

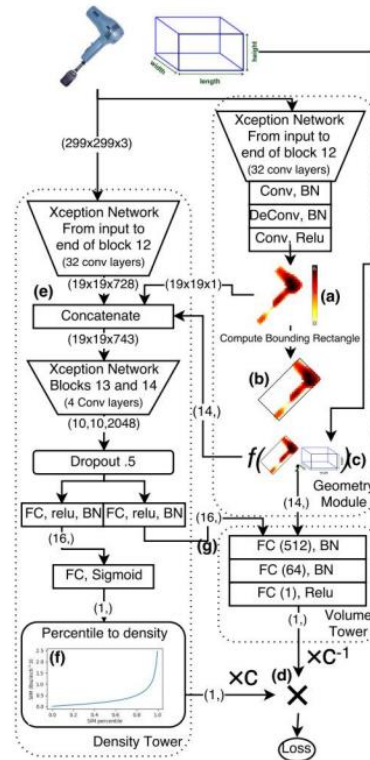


図 1 image2mass のモデル図。[3]より引用。

### 3. 提案

2. で述べた研究はいずれも画像のみから重量推定を行っているわけではなく、物体が外接する直方体のサイズ、もしくは物体の密度に関する情報も入力として必要としている。物体の画像に加えてサイズや密度などのような情報が入手できれば、より高い精度の重量推定が可能になると考えられるが、実際の現場ではこれらの情報は必ず入手できるとは限らない。人間の場合はそのような状況でも、画像のみからある程度の精度で重量を推定し現場での問題に対処できることから、本研究では単一画像のみから重量推定を行う手法を提案する。

<sup>†</sup> 東京工科大学 Tokyo University of Technology

提案する手法を具体的に解説する。本研究では、単一画像のみの入力から画像中の物体の重量推定を行い、重量のスカラ値を出力するような CNN のモデルを提案する。具体的には、CNN アーキテクチャのバックボーンの部分はそのままで、終端の部分、線形関数などを利用して少し変更することで、本来は画像分類などに利用されるような CNN モデルを利用して回帰を行う。図 2 は提案手法のモデル図である。

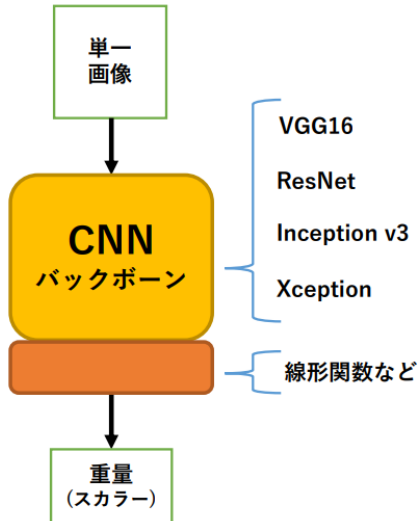


図 2 提案手法のモデル図

使用した CNN モデルは VGG16[8]、ResNet[9]、Inception v3[10]、Xception[11]である。

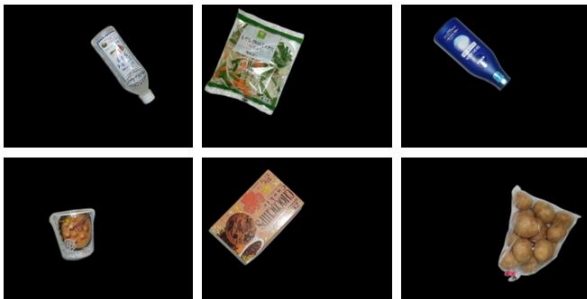


図 3 データセットに使用した画像の例

図 3 は実験に用いたデータセットの画像の例である。実験に使用したデータセットについて解説する。使用したデータセットは一般的なスーパーで売られている商品を撮影したものである。撮影する際には、1 枚の画像内には 1 つの商品だけが映る用に撮影した。また、商品の背景は黒一色に統一した。その際、撮影後に画像内の背景のノイズをプログラムで除去することで、より鮮明な黒色の背景にした。また、商品のサイズは 40cm×40cm×40cm の立方体に収まるサイズに限定した。また、それぞれの商品に対しておよそ 10 枚ずつ、位置、角度を変えながら、寝かせた状態で上から撮影した。商品の種類としては食品が多いが、一般的なスーパーやドラッグストアにあるような雑貨品も含まれている。表 1 に使用したデータセット内の商品のカテゴリーごとの種類数を示す。

#### 4. まとめと今後の課題

実験の結果、提案手法は人間による重量推定と比較しても遜色がなく、十分に実用的であると言える。

また、精度がさらに高くなることで、実際の現場でのロボットの作業効率が向上することが見込めるため、さらなる精度向上を目指す価値がある。

しかし、今回使用したデータセットはスーパーで販売されている商品に限定されているため、より一般的な物体に対する重量推定の精度を検証する余地があると言える。

表 1 商品のカテゴリーと種類数

カテゴリー	種類数
野菜・果物	20
生肉・生魚	12
袋入りお菓子・パンなど	23
箱入りお菓子・カップスープなど	16
飲料・缶詰・ジャムなど	27
日配品・その他食品	33
洗剤・ボディークリームなど	10
その他雑貨品	14

#### 参考文献

- [1] Ü. B. Şalvarci and U. E. Ayten, "Distance independent weight estimation of eggs from images using artificial neural networks," *2017 25th Signal Processing and Communications Applications Conference (SIU)*, Antalya, Turkey, pp. 1-4, doi: 10.1109/SIU.2017.7960637, (2017).
- [2] 福島 誠人, 竹長 慎太郎, 波部 齊, 井口 信和, "画像解析による植物の重さ推定の提案", 2020 年度情報処理学会関西支部 支部大会, G-50, (2020).
- [3] Trevor Standley, Ozan Sener, Dawn Chen, Silvio Savarese, "image2mass: Estimating the Mass of an Object from Its Image", *Proceedings of the 1st Annual Conference on Robot Learning*, PMLR 78:324-333, (2017).
- [4] J. M. L. Andrade and P. Moreno, "Improving the Estimation of Object mass from images," *2023 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Tomar, Portugal, pp. 199-206, doi: 10.1109/ICARSC58346.2023.10129573, (2023).
- [5] Xie, H., Yao, H., Zhang, S. et al. "Pix2Vox++: Multi-scale Context-aware 3D Object Reconstruction from Single and Multiple Images", *Int J Comput Vis* **128**, 2919–2935 (2020).
- [6] Dalai, Radhamadhab & Senapati, Kishore, "Weight Estimation through Image Analysis", *International Journal of Computer Trends and Technology*. 49. 71-75. 10.14445/22312803/IJCTT-V49P111, (2017).
- [7] R. Dalai and K. K. Senapati, "A Heuristic Grid Area Based Segmentation Approach for Weight Estimation of an Object from Image," *2018 4th International Conference for Convergence in Technology (I2CT)*, Mangalore, India, pp. 1-5, doi: 10.1109/I2CT42659.2018.9058247, (2018).
- [8] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [9] He, K., Zhang, X., Ren, S., & Sun, J. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, (2016).
- [10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. "Rethinking the inception architecture for computer vision." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818-2826, (2016).
- [11] Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251-1258, (2017).