

遮蔽のある不特定種類の物体に対する画像認識を用いた重量推定

Weight Estimation for Various Types of Occluded Objects Using Image Recognition

菅 シャフィウルハサン¹⁾
指導教員 青木 輝勝²⁾

- 1) 東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻
2) 東京工科大学 コンピュータサイエンス学部

ロボットが周囲の物体を操作する際に画像のみで重量を推定できれば作業効率が大幅に向上する。本研究では、画像のみで遮蔽のある不特定物体の重量を推定するため、GANによるハードサンプルマイニングと、セグメンテーションによって生成された遮蔽領域のマスク画像の利用の2つの手法を提案する。

画像認識, 重量推定, 遮蔽, GAN, セグメンテーション

1. 研究の背景と目的

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。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1つ存在しか存在しない[1]。また、そ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1つ存在する[2]。

また、画像を用いて物体を扱う場合には必ず発生する遮蔽の問題に対応した手法は存在しないため、本研究でそのような場合に対応した手法を提案する。

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[1]。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。

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[2]。Pix2Vox++は画像から画像内の物体の形

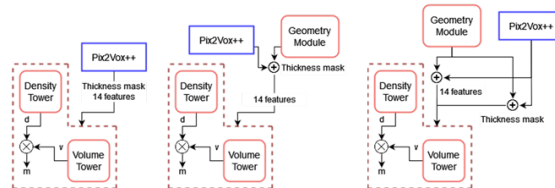


図1 [2]で提案された手法の例

状を推定することができる[3]。

Wangらは物体検出において、対象の物体に遮蔽や変形がある例に対しての精度向上の手法を提案した[4]。この研究では、学習中の物体検出器の中間層に対して、疑似的に遮蔽や変形を発生させることで、そのような難しい例に対して効果的に学習をさせた。

3. 提案

以下の提案手法①と提案手法②を提案する。また、本研究では既存研究の image2mass と同様に図2のような背景の切り抜かれた物体の画像を利用する。

提案手法①は、GANを用いて遮蔽のある物体の画像を生成し、その画像を用いて重量推定ネットワークに対してハードサンプルマイニングを行いながら学習をさせる手法である。



図2 提案手法①の入力画像と生成画像の例

ハードサンプルマイニングとは、学習過程でモデルがうまく認識できない、または誤分類しやすい「難しい」サンプルを特定し、それらを重点的に学習する戦略である。これによって、過学習の防止や精度の向上を行うことができる。

図2のように、入力画像に対して遮蔽のみを発生させるように学習されたGANを利用して画像を生成する。最終的な重量推定の損失は image2mass のみならず、GANの乱数生成器にもフィードバックされる。これによって、重量推定ネットワークの精度が高まるにつれてより難しい遮蔽を発生させる。

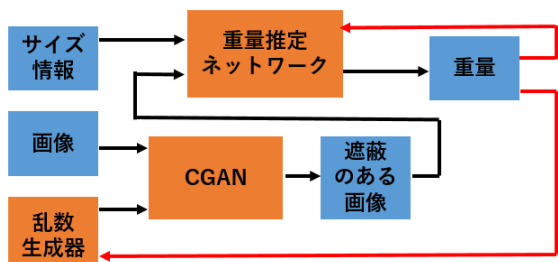


図3 提案手法①の処理の流れ(赤矢印は損失)

提案手法②は、遮蔽領域を推定することができる既存のセグメンテーションモデルを利用して、遮蔽領域を補完したマスク画像を生成し、それをチャンネル方向に結合した画像を重量推定ネットワークの入力画像として利用するというものである。

実験は image2mass を利用して行う。遮蔽に対応した既存の構造推定手法は多く存在するため、本研究では構造推定の改良は行わない[6]。

4. まとめと今後の課題

本研究では、遮蔽のある不特定種類の物体に対する画像認識による重量推定の精度向上の手法として、GANを用いたハードサンプルマイニングと遮蔽領域の推定を行ったマスク画像の利用という二つの提案を行った。これらの手法は重量推定の他にも多くのタスクに応用が可能だと考えられるため、今後のそのような方向性で研究を行いたい。

5. 参考文献

- [1] 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).
- [2] 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).
- [3] 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).
- [4] X. Wang, A. Shrivastava and A. Gupta, "A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 3039-3048, doi: 10.1109/CVPR.2017.324.
- [5] K. Chen, Y. Chen, C. Han, N. Sang, C. Gao and R. Wang, "Improving Person Re-Identification by Adaptive Hard Sample Mining," *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, 2018, pp. 1638-1642, doi: 10.1109/ICIP.2018.8451129.
- [6] Yang, Shuo, et al. "Single-view 3d object reconstruction from shape priors in memory." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.