Vol.2018-CVIM-210 No.50
2018/1/19

Multi-task CNNを用いた 単眼RGB画像からの超二次関数推定

八馬　遼[1,a)]　小篠　裕子[1,b)]　斎藤　英雄[1,c)]

概要：超二次関数は様々な基本形状を少ないパラメータ数で表現可能な形状表現手法であり，3 次元物体に当てはまる超二次関数パラメータを推定することで物体の形状を近似する技術はロボットの物体把持分野などで広く用いられている．超二次関数パラメータは同一の物体のカテゴリ間で類似する特性がある．本研究ではこの特性に着目し，超二次関数推定タスクと物体カテゴリ推定タスクを同時に学習させることで互いの精度向上を図る Multi-task CNN を超二次関数パラメータ推定に適用する．パラメータ推定は回帰問題とみなすことができるため，Regression CNN と Multi-task CNN を組み合わせた Multi-task Regression CNN を用いて超二次関数パラメータを推定する．実験では，超二次関数推定タスクと物体カテゴリ推定タスクのマルチタスク学習を超二次関数パラメータ推定に用いることの有効性を確かめた．

キーワード：超二次関数，マルチタスク学習，基本形状，CNN

Superquadric Parameter Prediction from a Single RGB Image by Multi-task CNN

Ryo Hachiuma[1,a)]　Yuko Ozasa[1,b)]　Hideo Saito[1,c)]

1. はじめに

3 次元物体形状を，円柱や直方体などの基本形状に近似する研究が数多く提案されている．特に，ロボットインタラクション分野において，3 次元物体の形状を基本形状で近似し，物体把持に用いる研究が注目されている．例えば，Harada ら [1] は，全ての把持対象物体を円柱で近似し，物体の把持位置を推定する手法を提案している．ロボットの物体把持に 3 次元基本形状近似を用いる研究には，全ての把持対象物体を円柱や直方体などの同一形状で近似する研究と，ひとつの基本形状に絞ることなく，物体に合わせて円柱や直方体，球などで形状を近似する研究とがある．後者の研究には，様々な基本形状を表現可能である超二次関数 [2] が広く使われている．図 1 に，超二次曲面によって

$(\varepsilon_1, \varepsilon_2) = (0.1, 0.1)$　$(\varepsilon_1, \varepsilon_2) = (0.1, 1.0)$　$(\varepsilon_1, \varepsilon_2) = (1.0, 1.0)$

図 1　超二次曲面例．

表現可能な基本形状の例を示す．

超二次関数は，スケールを表現するスケールパラメータと形状を表現する形状パラメータを用いて，様々な基本形状を表現可能とする関数であり，以下のように表現することができる．

$$F(x,y,z,\boldsymbol{q}) = \left\{ \left(\frac{x}{s_1}\right)^{\frac{2}{\varepsilon_2}} + \left(\frac{y}{s_2}\right)^{\frac{2}{\varepsilon_2}} \right\}^{\frac{\varepsilon_2}{\varepsilon_1}} + \left(\frac{z}{s_3}\right)^{\frac{2}{\varepsilon_1}}, \quad (1)$$

ここで，スケールパラメータは $\boldsymbol{s} = (s_1, s_2, s_3)$，形状パラメータは $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ である．$\boldsymbol{q} = (\boldsymbol{s}, \boldsymbol{\varepsilon})$ である．s_1, s_2, s_3 は，各 xyz 軸方向の超二次曲面の大きさを表し，$\varepsilon_1, \varepsilon_2$ は z 軸方向と x-y 平面の直角具合を表すパラメータを示す．3 次元物体

[1] 慶應義塾大学理工学部
Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, 223-8522, Japan
[a)] ryo-hachiuma@keio.jp
[b)] yuko.ozasa@keio.jp
[c)] hs@keio.jp

に当てはまる超二次関数のパラメータを推定する際は，対象となる 3 次元物体の点群を取得し，Levenberg-Marquardt アルゴリズム [3](LM) に基づく手法 [4] を用いてパラメータが推定されることがほとんどである．

超二次関数のパラメータは同一の物体カテゴリ間で似通う傾向があることが知られており，超二次関数を特徴ベクトルとみなし，物体カテゴリを推定する研究などが提案されている [5], [6]．本研究では，この事実に着目し，物体カテゴリ間で超二次関数が類似する特性を生かして，超二次関数のパラメータを推定したい．

2 つの相異なるタスクを同時に学習させることで，互いの精度向上を可能とするマルチタスク学習がある [7]．本研究では，超二次関数パラメータを推定するタスクと物体カテゴリを推定するタスクを同時に学習させるマルチタスク学習を超二次関数パラメータ推定に適用する．

物体カテゴリを推定する研究では，RGB 画像を入力とし，Convolutional Neural Network (CNN) を用いるものが主流となっている [8], [9]．本研究においても，RGB 画像を入力とし，マルチタスク学習と CNN を組み合わせたマルチタスク CNN(Multi-task CNN) を用いて超二次関数のパラメータを推定する．

本研究では，RGB 画像から超二次関数パラメータを直接推定する．同様の問題設定をしている関連研究として，人体関節の座標位置推定 [10] や，カメラの位置姿勢推定 [11] などがある．これらの研究は，画像から直接パラメータを推定する問題を回帰問題とみなし，Regression CNN を用いている．本研究では，Multi-task CNN と Regression CNN を組み合わせた Multi-task Regression CNN(MTR CNN) を用いる．

本論文では，超二次関数パラメータが物体カテゴリ間に似通う特性に着目し，物体カテゴリ推定と合わせて超二次関数パラメータを推定するマルチタスク学習を，超二次関数パラメータ推定に適用することの有効性を検証する．具体的には，物体の RGB 画像に対し，Multi-task Regression CNN を用いて超二次関数パラメータを推定する．実験では，マルチタスク学習の効果を検証するため，マルチタスク学習を行わずにパラメータを推定した場合をベースラインとし，推定精度を比較評価した．また，物体のカテゴリ推定において，未学習のカテゴリに属する物体に対してもパラメータ推定をし，未学習物体に対するパラメータ推定の頑健性を評価した．

2. 超二次関数パラメータ推定のための Multi-task Regression CNN

本論文で用いる Multi-task Regression CNN (MTR CNN) のアーキテクチャを図 2 に示す．入力は物体の RGB 画像 $\boldsymbol{x}$ であり，出力は超二次関数パラメータ $\boldsymbol{q}$ と物体のカテゴリ $\boldsymbol{y}$ である．図 2 において，conv は畳み込み層を，pool はプーリング層を，fc は全結合層を示し，各層にそのパラメータを示す．畳み込み層におけるパラメータは (フィルタサイズ，ストライド数)，プーリング層におけるパラメータはストライド数，全結合層におけるパラメータは出力サイズを示す．

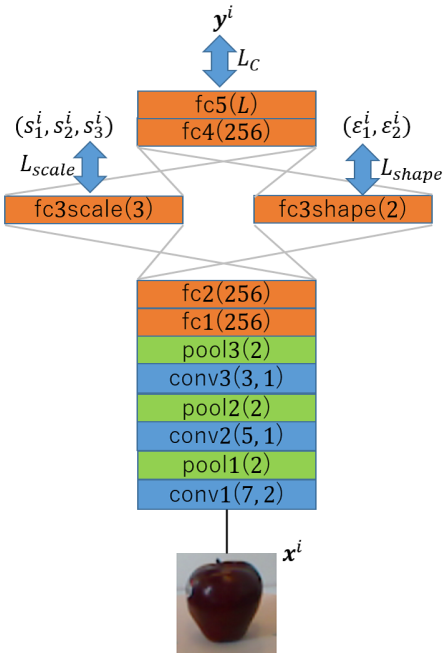

図 2 超二次関数パラメータ推定のための MTR CNN.

超二次関数パラメータ推定タスクとカテゴリ推定タスクを同時に行う MTR CNN の損失関数について説明する．超二次関数パラメータ推定タスクの損失関数を L_R，カテゴリ推定タスクの損失関数を L_C とすると，MTR CNN の損失関数 L_T は以下のようになる．

$$L_T = L_R(\hat{\boldsymbol{q}}, \boldsymbol{q}) + w_c L_C(\hat{\boldsymbol{y}}, \boldsymbol{y}), \tag{2}$$

ここで，$\hat{\boldsymbol{q}}, \boldsymbol{q}$ はそれぞれ超二次関数パラメータの真値と推定された値を，$\hat{\boldsymbol{y}}, \boldsymbol{y}$ はそれぞれカテゴリの真値と推定された値を表す．w_c は物体カテゴリ推定タスクの損失にかかる重みである．本章では，各タスクにおける損失関数を説明する．

2.1 超二次関数パラメータ推定タスクの損失関数

超二次関数パラメータ推定を回帰タスクとみなし，RGB 画像 $\boldsymbol{x}$ から，超二次関数パラメータ $\hat{\boldsymbol{q}}$ を推定する．形状パラメータ $\boldsymbol{\varepsilon}$ とスケールパラメータ $\boldsymbol{s}$ の特徴が異なるため，それぞれのパラメータで損失関数を算出する．形状パラメータの損失関数 L_{shape} は二乗平均平方根を用いて以下のように設定した．

$$L_{shape}(\hat{\boldsymbol{\varepsilon}}, \boldsymbol{\varepsilon}) = \sqrt{\frac{1}{d_1}\sum_i^{d_1} |\hat{\varepsilon}_i - \varepsilon_i|^2}, \tag{3}$$

ここで，$\boldsymbol{\varepsilon}$ は形状パラメータの真値，$\hat{\boldsymbol{\varepsilon}}$ は予測された形状パラメータであり，d_1 は形状パラメータの要素数である．同様にして，スケールパラメータの損失関数 L_{scale} を以下のように設定した．

$$L_{scale}(\hat{\boldsymbol{s}},\boldsymbol{s}) = \sqrt{\frac{1}{d_2}\sum_{i}^{d_2}|\hat{s}_i - s_i|^2}, \tag{4}$$

ここで，$\boldsymbol{s}$ はスケールパラメータの真値，$\hat{\boldsymbol{s}}$ は予測されたスケールパラメータであり，d_2 はスケールパラメータの要素数である．そして，超二次関数パラメータ推定の損失関数 $L_R(\hat{\boldsymbol{q}},\boldsymbol{q})$ は形状パラメータの損失関数 $L_{shape}(\hat{\boldsymbol{\varepsilon}},\boldsymbol{\varepsilon})$ とスケールパラメータの損失関数 $L_{scale}(\hat{\boldsymbol{s}},\boldsymbol{s})$ を用いて次のように表される．

$$L_R(\hat{\boldsymbol{q}},\boldsymbol{q}) = L_{shape}(\hat{\boldsymbol{\varepsilon}},\boldsymbol{\varepsilon}) + w_s L_{scale}(\hat{\boldsymbol{s}},\boldsymbol{s}), \tag{5}$$

ここで，w_s はスケールパラメータの損失関数にかかる重みである．

2.2 物体カテゴリ推定タスクの損失関数

物体のカテゴリ推定タスクを分類タスクとみなし，RGB 画像 $\boldsymbol{x}$ から物体のカテゴリ $\hat{\boldsymbol{y}}$ を推定する．物体カテゴリ推定の損失関数 $L_C(\hat{\boldsymbol{y}},\boldsymbol{y})$ は交差エントロピーを用いて次のように設定した．

$$L_C(\hat{\boldsymbol{y}},\boldsymbol{y}) = -\sum_{i}^{M}\hat{y}_i log(y_i), \tag{6}$$

ここで，M は推定するカテゴリ数を表す．

2.3 ネットワークの学習

損失関数には式 (2) を用いる．すべての畳み込み層と全結合層 (fc1, fc2, fc4) の出力には活性化関数として ReLU[12] が適用され，形状パラメータを出力する全結合層 (fc3shape) の出力には sigmoid 関数が，物体カテゴリを出力する全結合層 (fc5) の出力には，Softmax 関数が適用される．また，学習時には過学習を防ぐために最初の全結合層 (fc1) に dropout[13] を用いる．

3. 実験条件

3.1 データセット

本実験における評価のため，2 つのデータセットを構築した．1 つ目のデータセット (データセット A) は，超二次関数パラメータ推定の予測精度の評価のためのデータセットであり，2 つ目のデータセットはパラメータ推定の頑健性の評価のためのデータセット (データセット B) である．各データセットは，物体の RGB 画像とその物体の超二次関数パラメータの真値の組からなる．超二次関数パラメータの真値としては各物体の 3 次元点群から LM アルゴリズムにより推定された値を用いた．

本実験で用いたデータセットは RGB-D Object Dataset[14] をもとに構築した．RGB-D Object Dataset には 51 カテゴリの計 300 個の物体，各物体につき 600 個の RGBD 画像，3 次元点群が含まれる．各物体は様々なカメラ位置姿勢から撮影された．

RGB-D Object Dataset から各基本形状物体 (直方体，球，円柱) につき 3 つのカテゴリの物体を抽出し，データセット A を構築する．本モデルを評価には，Leave-one-object-out を用いる．これは物体の各カテゴリから，1 つの特定物体を各カテゴリからランダムに抽出しそれ以外の物体を学習に，その物体をテストに用いる．学習データは約 37,000 個，テストデータは約 6,700 個の RGB 画像と超二次関数パラメータの組から構成されている．

さらに，本モデルの頑健性を評価するため学習データ (データセット A) に含まれていないカテゴリの物体からなるデータセット B を構築した．データセット B に含まれる物体は，RGB-D object dataset の中から，2 つの基本形状，1 つの非基本形状物体を抽出した．

3.2 ベースライン手法

マルチタスク学習を用いた効果を検証するため，比較対象としてのベースライン手法として超二次関数パラメータのみを推定する Single Task Regression CNN(STR CNN) を用いた．このモデルのアーキテクチャは図 2 において，fc4, fc5 層を取り除いたものであり，損失関数には式 5 を用いた．

3.3 CNN の学習とハイパーパラメータ

入力される物体の RGB 画像は 149x149 にリサイズする．データセット A には 9 つのカテゴリの物体が含まれるため，図 2 における fc5 層の出力サイズは $M = 9$ となる．バッチサイズを 256 とし，畳み込み層と全結合層の重みは平均 0 標準偏差 0.1 の正規分布で初期化した．学習において dropout の割合は 0.5 とした．式 5 における重みを $w_s = 1.0$，式 2 における重みを $w_c = 10.0$ とし，学習率 0.0001 の Adam[15] を使用して損失関数を最適化する．

3.4 学習時とテスト時の損失

本論文で用いたモデルの損失の収束を確認するため，学習回数ごとの学習データの損失とテストデータの損失の変化を図 3.4 にまとめる．形状，スケールパラメータの損失をモデルごとに比較すると，STR CNN においては学習回数が進んでもテスト時の損失が減少しづらい傾向があるのに対し，MTR CNN においては減少傾向が確認され，より理想的に学習がなされていることが分かる．

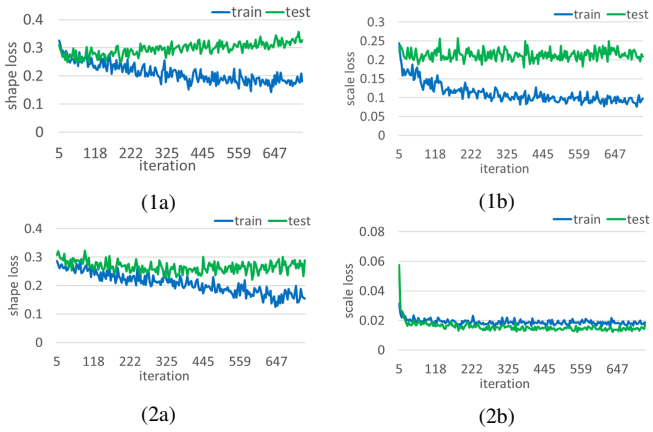

図 3 学習回数ごとの損失の変化 (a: 形状パラメータ，b: スケールパラメータ，1:STR CNN，2:MTR CNN).

4. 結果

4.1 パラメータ推定の精度評価

超二次関数パラメータ推定を定性的に評価するため，図 4 に MTR CNN を用いて推定された超二次曲面を示す．図 4 の上段は，物体の RGB 画像，下段は推定された超二次曲面である．また，各物体のに対し推定された形状パラメータは超二次曲面の図の下に，スケールパラメータは図中に示されている．例えば，apple の形状パラメータは $(\varepsilon_1, \varepsilon_2) = (0.686, 0.889)$，スケールパラメータは $(s_1, s_2, s_3) = (0.034, 0.033, 0.033)$ である．概ね直感に近い基本形状を推定できたことが分かる．

表 1 にカテゴリごとの真値との形状，スケールパラメータの誤差をまとめる．マルチタスク学習の効果を検証するため，STR CNN との誤差を比較した．形状，スケールパラメータの誤差はともにマルチタスク学習を用いることで，ほとんどのカテゴリにおいて誤差が小さくなったため，MTR CNN はより真値に近い超二次関数パラメータを推定したことが分かる．

データセット A は基本形状物体から構成されているため，物体の向きによる形状パラメータは小さいはずである．各物体は，様々な方向から撮影されているため，全ての形状パラメータ推定結果の標準誤差を算出することで，物体向き変化への頑健性を評価可能となる．MTR CNN と LM アルゴリズムにより推定された各形状パラメータの標準誤差は，MTR CNN を用いた場合は $(0.053, 0.102)$，LM アルゴリズムは $(0.175, 0.260)$ と，MTR CNN の方が誤差が小さく，物体向きに頑健な形状パラメータを推定したことが分かる．

4.2 パラメータ推定の頑健性

次に，未知学習カテゴリの物体 (データセット B) に対してマルチタスク学習を用いることの有効性を検証する．図

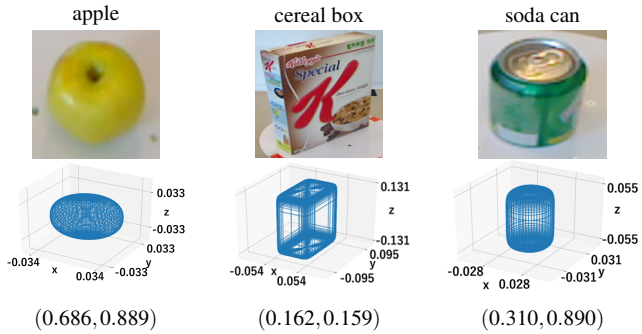

図 4 データセット A の物体に対して推定された超二次曲面の可視化例.

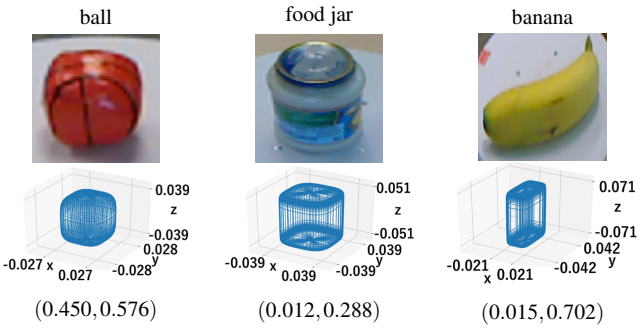

図 5 データセット B の物体に対して推定された超二次曲面の可視化例.

5 に推定された超二次曲面を可視化する．ball と food jar においては，概ねその物体の基本形状 (球，円柱) を成している．また，表 2 に LM アルゴリズムにより得られた形状，スケールパラメータとの誤差を示す．比較として，STR CNN との誤差を載せた．マルチタスク学習を用いることで，未学習カテゴリの物体に対してもパラメータの推定精度が向上した．

5. まとめ

本論文では，超二次関数パラメータに超二次関数パラメータ推定とカテゴリ推定のマルチタスク学習を用いることの有効性を検証した．本論文で構築した Multi-task Regression CNN では，超二次関数パラメータに加えて物体カテゴリも同時に学習することで，学習させたカテゴリの物体，学習させていないカテゴリの物体ともにパラメータの推定精度が向上することを示した．今後の課題として，超二次関数パラメータだけでなく，その位置姿勢も同時に学習することを目標とする．

謝辞 本研究の一部は，JST CREST の支援を受けたものである．

参考文献

[1] Harada, K., Nagata, K., Tsuji, T., Yamanobe, N., Nakamura, A. and Kawai, Y.: Probabilistic approach for object bin picking approximated by cylinders, *ICRA*, pp. 3742–3747 (2013).
[2] Barr, A. H.: Superquadrics and angle-preserving transformations, *IEEE Computer graphics and Applications*, Vol. 1,

誤差	手法	apple	cereal box	flashlight	food box	food can	kleenex	lime	soda can	tomato	平均
形状	STR CNN	0.254	0.065	0.216	0.048	0.256	0.064	0.349	0.200	0.289	0.193
	MTR CNN	0.199	0.053	0.131	0.038	0.227	0.058	0.339	0.189	0.241	0.164
スケール	STR CNN	0.009	0.033	0.013	0.016	0.015	0.241	0.007	0.010	0.008	0.015
	MTR CNN	0.006	0.023	0.012	0.014	0.014	0.179	0.009	0.008	0.006	0.012

表 1 推定された形状，スケールパラメータの真値との誤差.

category	method	scale error	shape error
ball	STR CNN	0.019	0.339
	MTR CNN	0.014	0.283
food jar	STR CNN	0.013	0.286
	MTR CNN	0.012	0.275
banana	STR CNN	0.021	0.247
	MTR CNN	0.017	0.241

表 2 LM アルゴリズムにより推定された形状パラメータとの誤差.

No. 1, pp. 11–23 (1981).

[3] Moré, J. J.: The Levenberg-Marquardt algorithm: implementation and theory, *Numerical analysis*, pp. 105–116 (1978).

[4] Solina, F. and Bajcsy, R.: Range image interpretation of mail pieces with superquadrics, *AAAI*, pp. 733–737 (1987).

[5] Xing, W., Liu, W. and Yuan, B.: Superquadric-based geons recognition utilizing support vector machines, *ICSP*, pp. 1264–1267 (online), DOI: 10.1109/ICOSP.2004.1441555 (2004).

[6] Hachiuma, R., Ozasa, Y. and Saito, H.: Primitive Shape Recognition via Superquadric Representation using Large Margin Nearest Neighbor Classifier, *VISAPP*, pp. 325–332 (2017).

[7] Caruana, R.: Multitask Learning, *Machine Learning*, Vol. 28, No. 1, pp. 41–75 (online), DOI: 10.1023/A:1007379606734 (1997).

[8] Wang, A., Lu, J., Cai, J., Cham, T. J. and Wang, G.: Large-Margin Multi-Modal Deep Learning for RGB-D Object Recognition, *IEEE Transactions on Multimedia*, Vol. 17, No. 11, pp. 1887–1898 (online), DOI: 10.1109/TMM.2015.2476655 (2015).

[9] Eitel, A., Springenberg, J. T., Spinello, L., Riedmiller, M. and Burgard, W.: Multimodal deep learning for robust RGB-D object recognition, *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 681–687 (online), DOI: 10.1109/IROS.2015.7353446 (2015).

[10] LI, S., Liu, Z. Q. and Chan, A. B.: Heterogeneous Multi-task Learning for Human Pose Estimation with Deep Convolutional Neural Network, *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 488–495 (online), DOI: 10.1109/CVPRW.2014.78 (2014).

[11] Massa, F., Marlet, R. and Aubry, M.: Crafting a multi-task CNN for viewpoint estimation, *CoRR*, Vol. abs/1609.03894 (online), available from ⟨http://arxiv.org/abs/1609.03894⟩ (2016).

[12] Nair, V. and Hinton, G. E.: Rectified linear units improve restricted boltzmann machines, *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814 (2010).

[13] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R.: Improving neural networks by preventing co-adaptation of feature detectors, *arXiv preprint arXiv:1207.0580* (2012).

[14] Lai, K., Bo, L., Ren, X. and Fox, D.: A large-scale hierarchical multi-view rgb-d object dataset, *ICRA*, pp. 1817–1824 (2011).

[15] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).