社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

畳み込みニューラルネットワークを用いた料理写真の魅力度推定

佐藤　陽昇† 道満　恵介††,† 平山　高嗣† 井手　一郎† 川西　康友†
出口　大輔†††,† 村瀬　洋†

† 名古屋大学 大学院情報学研究科 〒 464-8601 愛知県名古屋市千種区不老町
†† 中京大学 工学部 〒 470-0393 愛知県豊田市貝津町床立 101
††† 名古屋大学 情報戦略室 〒 464-8601 愛知県名古屋市千種区不老町
E-mail: †satoa@murase.is.i.nagoya-u.ac.jp, ††kdoman@sist.chukyo-u.ac.jp,
†††{hirayama,ide,kawanishi}@i.nagoya-u.ac.jp, ††††{ddeguchi,murase}@nagoya-u.jp

あらまし　我々は，料理を美味しそうに撮影するための支援技術として，料理が美味しそうに見える度合い「魅力度」を推定する手法を提案してきた．この手法では，魅力度付きの料理画像群から画像特徴を抽出し，回帰の枠組みにより未知の料理写真に対して魅力度を推定する．本報告では，新たなアプローチとして畳み込みニューラルネットワーク（CNN）を用いて魅力度推定器を構築した結果を述べる．具体的には，VGG16，ResNet50，Inception-v3 の学習済みモデルをそれぞれ転移学習した CNN を構築し，これを魅力度推定器として利用する．従来手法と比較した結果から，VGG16 の事前学習モデルに基づいた魅力度推定器が有効であることを確認した．
キーワード　料理写真，撮影支援，魅力度，畳み込みニューラルネットワーク

Estimating the attractiveness of a food photo using a Convolutional Neural Network

Akinori SATO†, Keisuke DOMAN††,†, Takatsugu HIRAYAMA†, Ichiro IDE†, Yasutomo KAWANISHI†, Daisuke DEGUCHI†††,†, and Hiroshi MURASE†

† Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan
†† School of Engineering, Chukyo University
100 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan
††† Information Strategy Office, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan
E-mail: †satoa@murase.is.i.nagoya-u.ac.jp, ††kdoman@sist.chukyo-u.ac.jp,
†††{hirayama,ide,kawanishi}@i.nagoya-u.ac.jp, ††††{ddeguchi,murase}@nagoya-u.jp

Abstract　We have previously proposed a method for estimating the attractiveness of a food photo in order to assist a user to shoot attractive food photos. In this method, image features were extracted from food photos with attractiveness scores, and the attractiveness score of an input food photo was estimated in a regression framework. In this report, we describe the result of constructing an attractiveness estimator using a convolutional neural network (CNN) as a new approach to this method. Specifically, CNNs which transfer-learned each of the pre-trained models of VGG16, ResNet50 and Inception-v3 is constructed and used as an attractiveness estimator. From the results compared with the previous method, we confirmed that the attractiveness estimator based on the VGG16 pre-trained model was effective.
Key words　Food photo, shooting support, attractiveness, convolutional neural networks

(a) 魅力に欠ける構図で撮影された料理写真　(b) 魅力的な構図で撮影された料理写真

図 1: 同一の料理を被写体とした料理写真の例.

1. はじめに

近年，料理レシピサイトや SNS の普及により Web 上への料理写真の投稿が増加している．Web 上に投稿される料理写真は美味しそうに撮影されていることが望ましい．しかし，Web サイトや SNS に投稿される料理写真は，同一の料理でも美味しそうに見える度合いが様々である．例えば，図 1 は同一の料理を撮影した料理写真であるが，図 1(a) よりも図 1(b) の方が，写真全体に占める大きさや構図の点で料理が美味しそうに撮影されている．このように，被写体の大きさや撮影構図，色合いなどの違いによって料理写真に対する美味しさの印象が変わると考えられる．そのため，美味しそうに見える料理写真を撮影するためには，これらの要因を考慮して適切な撮影方法を選ぶ必要がある．

しかし，非専門家にとって撮影方法を適切に決定することは必ずしも容易ではない．そのため，料理写真の撮影方法を推薦するシステムがあれば有用であると考えられる．そのようなシステムを実現するためには，まず撮影された料理が美味しそうに見える度合いを定量的に分析する必要がある．

我々は従来研究 [1] において，料理が美味しそうに見える度合いを「魅力度」と定義し，実験協力者が魅力度の一対比較を行う選好実験に基づいて料理画像に魅力度を付与したデータセット「NUFOOD 360×10」(注1)を構築した [2]．そして，料理画像の色彩調和，エッジの向きと強度，畳み込みニューラルネットワーク （CNN）の中間層から得られる特徴量である Deep Convolutional Activation Feature（DeCAF）[3]，主食材の大きさと位置と向きを特徴量として利用し，Random Regression Forest [4] を用いて回帰モデルを構築することで，料理写真の魅力度を推定する手法を提案した．この手法では一般的な審美性 [5] や撮影のノウハウ [6] を考慮してトップダウン的に特徴設計をしているが，食材構成を考慮していないため，これだけでは料理写真の魅力度を評価するための画像特徴として十分に表現することができていない [7].

これを解決する方法として，特徴をデータ駆動的に利用するために End-to-End に CNN を用いて魅力度推定器を構築する手法が考えられる．関連研究として，會下らは料理写真からのカロリー量推定において CNN を用いることで高精度な推定ができることを確認した [8]．また，Salvador らは料理写真からのレシピ推定において CNN を利用することの有効性を確認した [9].

(注1)：http://www.murase.is.i.nagoya-u.ac.jp/nufood/

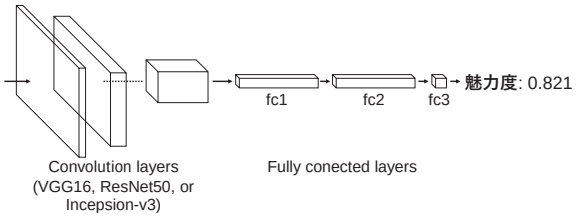


図 2: CNN のネットワークアーキテクチャ.

以上より，料理写真の魅力度推定においても，CNN の適用が有効であると考えられる．そこで，本報告では，魅力度推定器を CNN を用いて構築した結果について述べる．

以降，2 節で CNN を用いた料理写真の魅力度推定手法について説明する．次に 3 節で実験で使用する料理画像データセットについて述べ，4 節で提案手法の有効性を評価した実験について述べる．最後に 5 節で本報告をまとめる．

2. CNN を用いた料理写真の魅力度推定手法

本節では，提案する魅力度推定器を構成する CNN のネットワークアーキテクチャとその学習方法およびそれを用いた魅力度の推定方法について述べる．

2.1 ネットワークアーキテクチャ

大規模データセットを学習済みの CNN を，適用する問題に合わせたデータセットを用いて転移学習する手法 [10]〜[12] が広く利用されている．本報告においてもそれに倣い，転移学習に用いる学習済み CNN モデルとして，ImageNet の 1000 クラス分類タスクで学習された VGG16 [13]，ResNet50 [14]，Inception-v3 [15] を扱う．VGG16 は，単純に積層した畳み込み層が 13 層，全結合層が 2 層，出力層が 1 層のネットワークである．ResNet50 は，ある層の出力とそれよりも前の出力との残差を学習することにより，構造を深くすることを可能としたネットワークで 49 層の畳み込み層を持つ．Inception-v3 は Inception モジュールという並列した畳み込み層の組ごとに損失を伝搬させるネットワークである．

提案手法のネットワークアーキテクチャを図 2 に示す．畳み込み層は VGG16，ResNet50，Inception-v3 いずれかの出力層側にある全結合層を除いた層からなり，それ以降に，256 次元の全結合層である fc1 と fc2 と単一のユニットで構成される出力層である fc3 を持つ．なお，fc1 の入力次元数は基とする事前学習モデルの出力層側にある全結合層前の出力次元数により決まる．fc1 と fc2 の活性化関数には ReLU を使用し，Batch Normalization [16] を適用する．また，魅力度を [0,1] で出力するために，fc3 にシグモイド関数を適用する．なお，入力となる画像サイズは 244×244[px] である．

2.2 魅力度推定器の学習

提案手法では，3 節で紹介するデータセット内の料理画像を 244×244[px] にリサイズしたものを学習画像，その料理画像の魅力度を教師信号として 2.1 節のネットワークを転移学習する．

(a) 鰹のたたき　(b) カレーライス　(c) 鰻丼　(d) ビーフシチュー　(e) ハンバーグ

(f) 天丼　(g) カツ丼　(h) 鉄火丼　(i) チーズバーガー　(j) フィッシュバーガー

図 3: 「NUFOOD 360×10」 [1], [2] に含まれる各料理.

ただし，学習はモデル全体ではなく，全結合層 fc1 と fc2 のパラメータのみ行う．また，損失関数は出力値と教師信号との絶対誤差の平均とする．

2.3 魅力度推定器による推定

まず，入力した料理画像を 244×244[px] にリサイズする．そして，2.2 節で学習した魅力度推定器を用いて魅力度を算出する．

3. 魅力度付き料理画像データセット

本節では，実験用データセットとして用いた魅力度付き料理画像データセット「NUFOOD 360×10」 [1], [2] について紹介する．このデータセットは，同一の料理について仰角と回転角を変更して 36 方向から撮影した料理画像群に対する選好実験の結果に対して，Thurstone の一対比較法 [17] を適用して算出した魅力度を付与したものである．

以下，このデータセットの構築方法を紹介する．

3.1 対象料理

色合いや立体感の違いを考慮して 10 種類の料理が選ばれた．具体的には，鰹のたたき，カレーライス，鰻丼，ビーフシチュー，ハンバーグ，天丼，カツ丼，鉄火丼，チーズバーガー，フィッシュバーガーである．撮影の利便性と再現性の点から，時間経過に伴う状態の変化や盛り付けの変化が生じない食品サンプルが用いられた．

3.2 料理画像群

図 3 に各料理の画像群の抜粋を示す．撮影角度として設定された仰角は，撮影装置の回転皿と同じ平面を仰角 0 度とし，その面を基準に 30，60，90 度であった．回転角は，料理のある方向を基準として，その方向から右回りに 30 度刻みに 330 度までの角度であった．

3.3 魅力度の付与

高橋らは，Thurstone の一対比較法 [17] により，料理画像群の各料理画像に魅力度を付与した．Thurstone の一対比較法は官能検査の 1 つであり，対比較結果に基づいて複数の試料の感覚値を間隔尺度化するものである．

具体的には，まず，各料理画像群の料理画像 36 枚に対して，異なる 2 枚の組み合わせ ${}_{36}C_2 = 630$ 通りの料理画像対を生成した．次に，全ての組み合わせに対して各々 3 人または 4 人の実験協力者から回答が得られるように一対比較による選好実験を行った．実験協力者は，料理画像対が左右に提示され，「美味しく見える方はどちらか」という設問に対して，「左」または「右」と回答し，判断できない場合には「分からない」と回答した．本データセットを構築した際には，20 代の男女延べ 28 名の実験協力者から料理ごとに 2,015 件の選好結果を得た．そして，得られた選好結果に対して，間隔尺度値を求めた．最後に，間隔尺度値を [0,1] に正規化し，料理画像の魅力度とした．魅力度が 1 に近いほど魅力度が相対的に高い画像である．この処理を各料理に適用することで，魅力度付き料理画像データセットを構築した．

3.4 一対比較により付与された魅力度

図 4 に，付与された魅力度により順位付けされた各料理画像群の抜粋を示す．全ての料理に共通して，多種の食材が見えるように撮影された画像や，その料理を特徴づける食材が手前に撮影された画像の魅力度は高い傾向が見られた．また，見えが似ているビーフシチューとハンバーグは，全体的に撮影構図とその魅力度の関係が似ていた．一方，料理全体や主食材が見える領域が少ない角度で撮影された画像や，立体感が感じられない画像の魅力度は低い傾向が見られた．

4. 評価実験

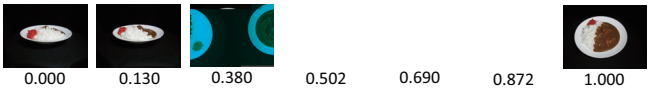
本節では，実験により VGG16 [13], ResNet50 [14], Incepiton-v3 [15] それぞれを転移学習した提案手法の有効性について評価した結果について述べる．

4.1 実験条件

比較手法には，高橋らが提案した料理画像の色彩調和，エッジの向きと強度，DeCAF [3]，主食材の大きさと位置と向きを特徴量として利用し，Random Regression Forest [4] を用いて回帰モデルを構築する手法 [1] を用いた．事前処理として特徴量の各次元を [0,1] に正規化した．

推定器の構築および評価は，3 節で述べたデータセット内の各料理画像群で学習した推定器ごとに Leave-One-Out 法を用

(a) 鰹のたたき



(b) カレーライス

(c) 鰻丼

(d) ビーフシチュー

(e) ハンバーグ

(f) 天丼

(g) カツ丼

(h) 鉄火丼

(i) チーズバーガー

(j) フィッシュバーガー

図 4: 選好実験により付与された魅力度付き料理画像群（抜粋）[1], [2].

表 1: 各手法による魅力度の推定誤差（表中の数値は MAE [0,1] を示し，下線に太文字は各料理において推定誤差が最小の値，太文字は 2 番目に小さい値を示す）.

料理名	手法			
	比較 [1]	VGG16	ResNet50	Inception-v3
鰹のたたき	**0.128**	**0.127**	0.144	0.161
カレーライス	**0.085**	**0.083**	0.125	0.094
鰻丼	**0.070**	**0.070**	**0.068**	0.071
ビーフシチュー	**0.087**	**0.085**	0.128	0.160
ハンバーグ	**0.100**	**0.091**	0.109	0.094
天丼	**0.132**	**0.124**	0.135	**0.124**
カツ丼	**0.099**	**0.122**	0.128	0.148
鉄火丼	0.054	**0.042**	0.050	**0.043**
チーズバーガー	**0.072**	0.086	0.095	**0.082**
フィッシュバーガー	**0.081**	**0.093**	0.120	0.147
平均	**0.091**	**0.092**	0.110	0.117

いた．提案手法の CNN の学習には Deep Learning 用フレームワークである Keras(注2)を使用し，そのバックエンドとして TensorFlow(注3)を使用した．また，エポック数は 100，バッチサイズは 18 とし，最適化手法に Adam を使用した．比較手法の Random Regression Forest の学習には，scikit-learn ライブラリ [18] の RandomForestRegressor を利用し，パラメータは random_state = 2，n_estimators = 150 とした．評価指標は，Thurstone の一対比較法により算出した魅力度と，各手法により推定された魅力度との平均絶対誤差（MAE：Mean Absolute Error）とした．

4.2 実験結果

提案手法の有効性を評価した結果を表 1 に示す．ResNet50 と Inception-v3 はそれぞれ鰻丼，天丼において推定誤差が最小となったが，平均的には比較手法より推定誤差が大きくなった．VGG16 は 6 つの料理において推定誤差が最小となり，平均的に比較手法との差は僅かであった．以上より VGG16 を転移学習する提案手法が有効であると確認できた．

4.3 考察

VGG16 は ResNet50 と Inception-v3 より推定精度が良好だった．このことから，料理写真の魅力度推定における CNN の構造は単純な積層構造が有効である可能性が示唆される．それでも，従来手法の推定精度を上回ることはなかった．しかし，魅力度推定のための画像特徴抽出を手細工で設計する必要がない点において提案手法は優れているといえる．

提案手法では，ImageNet の 1000 クラス分類タスクにおける事前学習モデルを転移学習した．しかし，この事前学習モデルは，汎用的な画像分類に適したものであり，料理画像における特徴を十分に捉えられていない可能性がある．これを解決するために，事前学習に用いるデータセットを料理画像のみとして，料理カテゴリ分類タスクにより事前学習したものを転移学習する手法が考えられる．料理画像のみの大規模データセットとして，

(注2)：https://keras.io/
(注3)：https://www.tensorflow.org/

100 種類の料理カテゴリをラベル付けした「UEC-FOOD100」[19](注4)が挙げられる．このデータセットを用いて事前学習した CNN を転移学習することで，料理画像の特徴を十分に捉えた魅力度推定器を構築することができる．

5. ま と め

料理をおいしそうに撮影するための支援を目的とし，畳み込みニューラルネットワーク（CNN）を用いて料理写真の魅力度を推定する手法を提案し，評価した．評価実験により，VGG16 を転移学習した CNN を用いる提案手法の有効性を確認した．

今後は，大規模な料理写真データセットを用いて事前学習したモデルを利用することを検討する．また，様々な料理カテゴリに渡った汎用的な推定器の構築や撮影支援システムへの応用についても検討していく．

謝辞 本研究の一部は，科研費および MSR-CORE12 による．

文 献

[1] K. Takahashi, K. Doman, Y. Kawanishi, T. Hirayama, I. Ide, D. Deguchi, and H. Murase, “Estimation of the attractiveness of food photography focusing on main ingredients,” Proc. 9th Workshop on Cooking and Eating Activities (CEA2017) in conjunction with IJCAI2017, pp.1–6, 2017.

[2] 井手一郎，高橋和馬，道満恵介，川西康友，平山高嗣，出口大輔，村瀬 洋，“視点に応じた魅力度が付与された料理画像データセット，” 平成 29 年度電気・電子・情報関係学会東海支部連合大会，Po1-25，2017．

[3] J. Donahue, Y. Jia, O. Vinvals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, “Decaf: A deep convolutional activation feature for generic visual ecognition,” Proc. 31st Intl. Conf. on Machine Learning, pp.647–655, 2014.

[4] L. Andy and M. Wiener, “Classification and regression by random forest,” R News, vol.2, no.3, pp.18–20, 2002.

[5] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, “Aesthetic quality classification of photographs based on color harmony,” Proc. 2011 IEEE Conf. on Computer Vision and Pattern Recognition, pp.33–40, 2011.

[6] 佐藤 朗，もっとおいしく撮れる！お料理写真 10 のコツ，株式会社青春出版社，2012．

[7] 服部竜実，道満恵介，井手一郎，目加田慶人，“料理写真の魅力度を推定する際の画像特徴に関する定量分析—食材構成の理解が魅力度に及ぼす影響—，” 信学技報，MVE2017-22，2017．

[8] 曾下拓実，柳井啓司，“食事レシピ情報を利用した食事画像からのカロリー量推定，” 情処学研報，CVIM，vol.2017-CVIM-207-13，2017．

[9] A. Salvador, N. Hynes, Y. Aytar, J. Marin, F. Ofli, I. Weber, and A. Torralba, “Learning cross-modal embeddings for cooking recipes and food images,” Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition, pp.3068–3076, 2017.

[10] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, “Semantic image segmentation via deep parsing network,” Proc. 2015 IEEE Int. Conf.c on Computer Vision, pp.1377–1385, 2015.

[11] V. Gajarla and A. Gupta, “Emotion detection and sentiment analysis of images,” Georgia Institute of Technology, 2015.

[12] C. Szegedy, S. Reed, D. Erhan, D. Anguelov, and S. Ioffe, “Scalable, high-quality object detection,” arXiv preprint arXiv:1412.1441, 2014.

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

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition, pp.770–778, 2016.

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition, pp.2818–2826, 2016.

[16] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” Proc. 32nd Int. Conf. on Machine Learning, pp.448–456, 2015.

[17] L.L. Thurstone, “Psychophysical analysis,” American J. of Psychology, vol.38, no.3, pp.368–389, 1927.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Aesthetic quality classification of photographs based on color harmonscikit-learn machine learning in python,” J. of Machine Learning Research, vol.12, pp.2825–2830, 2011.

[19] Y. Kawano and K. Yanai, “Foodcam: A real-time food recognition system on a smartphone,” Multimedia Tools and Applications, vol.74, no.14, pp.5263–5287, 2015.

(注4)：http://foodcam.mobi/dataset100.html