一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
IMQ2021-68,IE2021-130,MVE2021-97(2022-03)

料理写真の魅力度推定に適したマルチタスク学習の検討

宮崎　光明†　　道満　恵介†　　井手　一郎††　　目加田慶人†

† 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 大学院情報学研究科　〒464–8601 愛知県名古屋市千種区不老町
E-mail: †miyazaki.m@md.sist.chukyo-u.ac.jp, {kdoman,y-mekada}@sist.chukyo-u.ac.jp
††ide@i.nagoya-u.ac.jp

あらまし　料理写真の魅力度推定に適したマルチタスク学習について検討した結果を報告する．これまで我々は，小規模データを有効活用することによって料理写真の魅力度推定精度を向上させる方法を検討してきた．本研究では推定精度を向上させる新たな方法としてマルチタスク学習による魅力度推定手法を提案する．マルチタスク学習とは，一つのモデルの中で複数のタスクを同時に学習する方法である．関連したタスクを同時に解くことにより，個別のモデルで解くよりもモデルの汎化性能が向上することが知られている．そのため，料理写真の魅力度推定においても，魅力度推定に関連したタスクを同時に解くことにより，魅力度推定精度の向上が期待できる．本報告では魅力度推定をメインタスクとし，魅力度推定に関連するタスクをサブタスクとしたマルチタスク学習を提案し，実験によりその有効性を確認する．
キーワード　料理写真，魅力度推定，マルチタスク学習，回帰分析

A study on multi-task learning for the attractiveness estimation of food photography

Mitsuaki MIYAZAKI†, Keisuke DOMAN†, Ichiro IDE††, and Yoshito MEKADA†

† School of Engineering, Chukyo University
101 Tokodachi, Kaizu-cho, Toyota, Aichi, 470–0393 Japan
†† Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464–8601 Japan
E-mail: †miyazaki.m@md.sist.chukyo-u.ac.jp, {kdoman,y-mekada}@sist.chukyo-u.ac.jp
††ide@i.nagoya-u.ac.jp

Abstract　We report the results of a study on multi-task learning for the accuracy improvement of attractiveness estimation of food photography. We have been studying a way of improving the estimation accuracy by effectively using a small-scale image dataset. This paper proposes another approach based on multi-task learning for attractiveness estimation. Multi-task learning is a method that simultaneously learns multiple tasks within one model. Solving a task of interest together with its related tasks can improve the generalization performance of a trained model compared to solving each task independently. Thus, we expect that such a multi-task approach is effective for the attractiveness estimation of food photography. This report proposes a multi-task learning method that simultaneously solves multiple problems including attractiveness estimation as the main task and related estimation/classification problems as subtasks, and also confirm its effectiveness through evaluation experiments.
Key words　Food photography, attractiveness estimation, multi-task learning, regression analysis

1. ま え が き

料理レシピサイトや SNS への料理写真の投稿機会が増えている．このような 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．例えば，図 1 は同一の料理を撮影した写真であるが，図 1a よりも図 1b の方が，被写体の大きさ，ぼかし，撮影角度等の観点で料理が美味しそうに撮影されている．料理写真を美味しそうに撮影することは必ずしも

(a) 魅力的でない料理写真

(b) 魅力的な料理写真

図 1: 魅力度が異なる料理写真の例

Fig. 1 Example of food photos with different attractiveness.

容易ではなく，魅力的な料理写真の撮影を支援するシステムがあれば有用である．本研究では，料理写真が美味しそうに見える度合いを「魅力度」と定義し，これを推定する技術の開発を目的とする．

これまで，佐藤らは画像選好実験時の選好者の視線情報を CNN による画像特徴抽出の際の重みとして用いることで，人の視線情報を考慮した画像特徴の設計を行った [1]．また，服部らは魅力度に影響を及ぼさない範囲で画像変換を適用することで，画像とその魅力度の組を生成する Data Augmentation によって，小規模データの拡充を行った [2]．

これらはいずれも，魅力度推定手法の精度向上を目指すものであった．これに対して本研究では，魅力度推定精度の向上のための別の方法として，マルチタスク学習を用いた魅力度推定手法を検討する．マルチタスク学習では各タスクに関連がある場合，同時に学習することによってタスク間に共通する特徴が学習され，モデルの汎化性能が向上することが知られている [3]．本報告では，魅力度推定をメインタスクとし，魅力度と関連するタスクをサブタスクとしたマルチタスク学習を提案する．

2. 料理写真の魅力度推定手法

魅力度推定手法は，図 2 に示すように，学習と推定の 2 段階で構成される．学習段階では，予め魅力度が付与された料理画像を用いて，各タスクの推定が可能な一つの推定器を構築する．推定段階では，学習段階で構築された推定器を用いて，入力された料理画像に対する各タスクの推定結果を出力する．また，本研究では推定器の汎化性能向上のためにマルチタスク学習に加え，服部らの Data Augmentation 手法 [2] を適用する．これは元画像と見え方の変化が知覚できない範囲で画像生成することによって，生成した画像に元画像と同じ魅力度を付与する方法である．画像変換は回転，拡縮，平行移動，ランダムノイズ付与の 4 種類を採用する．

2.1 マルチタスク学習

提案手法では図 3 のように，VGG16 [4] の最終畳み込み層までネットワークを共有し，その後，各タスクの全結合層へと入力して推定する形のマルチタスク学習を利用する．なお，マルチタスク学習ではサブタスクを複数設定することも可能であり，これにより，多くの特徴を考慮した学習ができ，モデルの汎化性能が向上する可能性がある．そのため，複数のサブタスクを

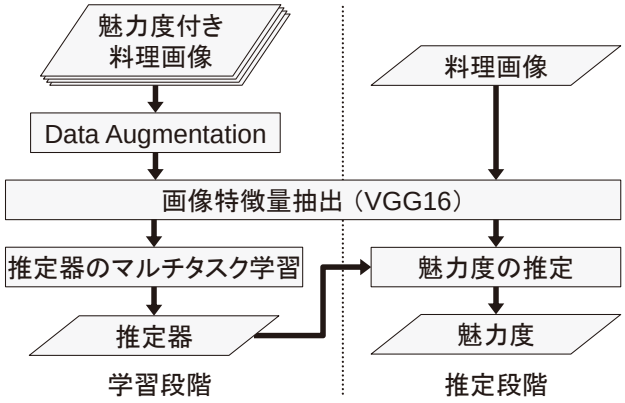


図 2: 魅力度推定の処理手順

Fig. 2 Process-flow of the estimation method.

設定したマルチタスク学習も検討する．マルチタスク学習ではメインタスク以外にもサブタスクの損失関数を定義する必要がある．本研究では式 (1) のように，メインタスクとサブタスクの損失関数に重みづけすることによって各タスクの重要度を決定する．

$$
\begin{aligned}
L &= \alpha L_{\mathrm{main}} + L_{\mathrm{sub}} \\
L_{\mathrm{sub}} &= \sum_{i=1}^{N} \beta_i L_i \\
\alpha &+ \sum_{i=1}^{N} \beta_i = 1.0
\end{aligned}
\tag{1}
$$

ここで，L_{main}，L_{sub}，α，β_i，N はそれぞれ，メインタスクの損失関数，サブタスクの損失関数，メインタスクの損失関数の重み，サブタスクの損失関数の重み，サブタスク数を表しており，$\alpha, \beta \geqq 0$ とする．なお本研究では，魅力度推定精度の向上を目的としているため，$\alpha > \beta_i$ となるように設定する．サブタスクは本研究で使用している魅力度付き料理画像データセット NU FOOD 360x10(注1)の構成を考慮して設定する．NU FOOD 360x10 では，仰角と回転角を変化させて 10 種類の料理を 36 方向から撮影されており，料理種毎に 0〜1 の範囲に正規化された魅力度が各画像に対して付与されている．これにより，サブタスクとして料理の種類分類，魅力度のクラス分類，仰角のクラス分類，回転角の角度差推定を行うタスクを設定する．以降，各サブタスクについて詳述する．

2.1.1 料理の種類分類

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の 10 種類の料理を分類する．本サブタスクは，料理の種類を分別する領域は，その料理を表す上で重要な領域であり，魅力度推定においても有効であると考えたため設定する．

(注1)：NU FOOD 360x10: https://www.cs.is.i.nagoya-u.ac.jp/opensource/nufood/

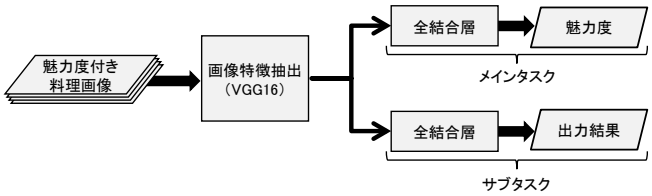


図 3: マルチタスク学習のネットワークモデル
Fig. 3　Network model of multi-task learning.

(a) 上位
魅力度：1.000

(b) 中位
魅力度：0.518

(c) 下位
魅力度：0.036

図 4: 魅力度の 3 クラスの例
Fig. 4　Example of three classes of attractiveness.

(a) 仰角：30°
魅力度：0.338

(b) 仰角：60°
魅力度：1.000

(c) 仰角：90°
魅力度：0.735

図 5: 仰角による魅力度の違い
Fig. 5　Difference in the attractiveness by elevation angles.

2.1.2　魅力度のクラス分類

0〜1 で付与されている魅力度を上位，中位，下位の 3 クラスに分類する．図 4 は，魅力度の 3 クラスの例を表したものであるが，魅力度のクラスによって料理の見えが大きく異なることが分かる．そのため，本サブタスクは，魅力度の大きさを区別する領域が魅力度推定においても有効であると考えたため設定する．また，本研究では魅力度の上位を魅力度 0.7 以上，中位を 0.3〜0.7，下位を 0.3 以下と実験的に定める．

2.1.3　仰角のクラス分類

30°，60°，90° の 3 種類の仰角を分類する．図 5 は，同一の回転角について異なる仰角で撮影した料理写真の例とその魅力度を示したものであるが，仰角が異なれば料理の見えや魅力度が異なることが分かる．そのため，本サブタスクを設定する．

2.1.4　回転角の角度差推定

12 種類の回転角（0°, 30°, . . . , 330°）の角度差を推定する．図 6 は，同一の仰角に異なる回転角で撮影した料理写真の例とその魅力度を示したものであるが，回転角が異なれば料理の見えや魅力度が異なることが分かる．そのため，本サブタスクを設定する．具体的には，魅力度が一番高い画像の回転角と他の画像の回転角の差分が最大値が 1 となるように正規化した値を回転角の角度差とし，これを推定する．また，角度差の推定は

(a) 回転角：60°
魅力度：1.000

(b) 回転角：150°
魅力度：0.398

(c) 回転角：240°
魅力度：0.000

図 6: 回転角による魅力度の例
Fig. 6　Difference in the attractiveness by rotation angles.

表 1: サブタスク数毎の損失関数の重み
Table 1　Loss function weights for each number of subtasks.

重み	サブタスク数（N）			
	1	2	3	4
α	0.70			Optuna [5] によって最適化（表 2）
β_i	0.30	0.15	0.10	

表 2: 最適化された損失関数の重み
Table 2　Optimized loss function weights.

タスク	メインタスク（α）	料理の種類分類（β_1）	魅力度のクラス分類（β_2）	仰角のクラス分類（β_3）	回転角の角度差推定（β_4）
重み	0.378	0.151	0.196	0.062	0.211

回帰分析によって行う．

3.　実　　験

マルチタスク学習を利用する提案手法の有効性を検証するため，2 節で述べたサブタスクを利用したマルチタスク学習が魅力度推定に及ぼす影響を定量的に分析した．以降，実験の方法および結果を述べ，考察する．

3.1　実 験 方 法

本実験では，マルチタスクモデルに対して層化 36 分割交差検証を適用し，魅力度の推定値と目標値の平均絶対誤差（MAE；Mean Absolute Error）を評価する．層化 36 分割交差検証とは，各クラスの割合が均一になるようにデータを分割して評価する方法である．本実験では，10 種類の料理から 1 枚ずつテスト画像を無作為にサンプリングし，残りの画像を全て学習画像とする．なお，学習画像に対して 2 節で述べた Data Augmentation [2] を適用することによって，元画像を含め計 14,350 枚の学習画像を用意する．次に，マルチタスク学習の条件としてサブタスク数毎の損失関数を表 1 のように定義する．ここで Optuna [5] とは，Bayesian 最適化によってハイパパラメータを求める Python ライブラリである．その結果を表 2 に示す．比較にはマルチタスク学習を行わないシングルタスクモデル（メインタスクのみでサブタスクなし）を用いる．各モデルは 30 epoch 学習し，回帰タスクと分類タスクの損失関数はそれぞれ MAE と Categorical Cross Entropy を用いる．また，事前に ImageNet [6] で学習済みの VGG16 を用いて転移学習する．

表 3: 各モデルの魅力度推定精度（MAE）
Table 3　Estimation accuracy (MAE) for each model.

料理	シングルタスクモデル	マルチタスクモデル			
		サブタスク			
		料理の種類分類	魅力度のクラス分類	仰角のクラス分類	回転角の角度差推定
鰹のたたき	0.169	0.154	0.167	0.158	0.169
カレーライス	0.111	0.114	0.119	0.107	0.111
鰻丼	0.087	0.091	0.095	0.084	0.084
ビーフシチュー	0.088	0.098	0.109	0.105	0.087
ハンバーグ	0.118	0.120	0.137	0.141	0.122
天丼	0.104	0.109	0.118	0.130	0.106
カツ丼	0.164	0.155	0.164	0.169	0.163
鉄火丼	0.067	0.063	0.053	0.052	0.067
チーズバーガー	0.103	0.102	0.103	0.116	0.105
フィッシュバーガー	0.112	0.080	0.100	0.094	0.116
平均	0.112	0.109	0.117	0.116	0.113

表 4: サブタスク数毎の最適なサブタスクの組み合わせ
Table 4　The best combination for each number of subtasks.

サブタスク数（N）	マルチタスクモデル			
	サブタスク			
	料理の種類分類	魅力度のクラス分類	仰角のクラス分類	回転角の角度差推定
1 種類	✓			
2 種類	✓			✓
3 種類	✓	✓		✓
4 種類	✓	✓	✓	✓

表 5: サブタスク数毎の最適な組み合わせの推定精度（MAE）
Table 5　Estimation accuracy(MAE) of the best combination for each number of subtasks.

料理	サブタスク数（N）			
	1	2	3	4
鰹のたたき	0.154	0.161	0.158	0.162
カレーライス	0.114	0.114	0.116	0.121
鰻丼	0.091	0.095	0.093	0.097
ビーフシチュー	0.098	0.097	0.098	0.090
ハンバーグ	0.120	0.120	0.115	0.112
天丼	0.109	0.109	0.104	0.104
カツ丼	0.155	0.154	0.156	0.159
鉄火丼	0.063	0.065	0.062	0.062
チーズバーガー	0.102	0.097	0.101	0.106
フィッシュバーガー	0.080	0.081	0.077	0.086
平均	0.109	0.109	0.108	0.110

3.2 実験結果

まず，単一のサブタスクを用いたマルチタスク学習の結果を表 3 に示す．同表の下線は料理種毎に推定精度が最も向上したモデルを表している．シングルタスクにおける MAE の平均値は 0.112 であるのに対して，料理の種類分類をサブタスクとして加えることで，MAE の平均値は 0.109 と推定精度が高くなった．続いて，サブタスク数毎の最適な組み合わせとそのときの魅力度推定精度をそれぞれ表 4 と表 5 に示す．表 5 の全ての組み合わせにおいてシングルタスクよりも平均して推定精度が高くなった．以上のことからマルチタスク学習の有効性を確認した．

3.3 考　　察

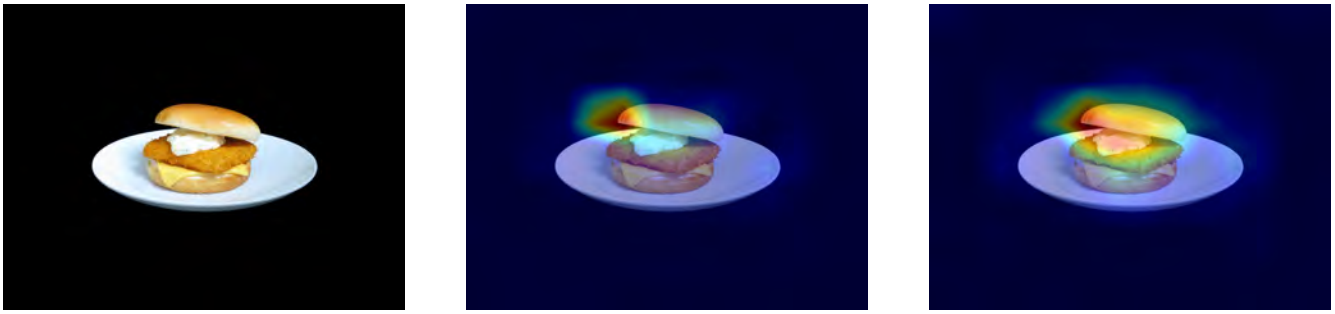
提案手法におけるマルチタスク学習が魅力度推定精度に及ぼす影響について考察する．

3.3.1　単一のサブタスクの利用が推定精度に及ぼす影響

料理の種類分類をサブタスクとしたマルチタスクモデルではシングルタスクモデルより平均して小さい推定誤差を示した．これは，料理の種類を分別する領域が魅力度推定に有効な領域と一致していたことによりモデルの汎化性能が向上したためと考えられる．特にフィッシュバーガーでは全てのモデルと比較して最も小さい推定誤差を示した．この要因を考察するためにシングルタスクモデルと料理の種類分類をサブタスクとして加えたマルチタスクモデルでフィッシュバーガーの画像特徴を Grad-CAM [7] によって可視化した．その結果を図 7 に示す．同図からフィッシュバーガーについてはシングルタスクモデルではバンズの領域のみが注目されているが，マルチタスクモデルではフィッシュフライやタルタルソースの領域も注目されていることがわかる．これによって，人間が魅力度を判断する際に注目していると考えられる，より重要な領域が抽出されたと考えられる．また，料理の種類毎に結果を比較したとき，ハンバーグと天丼以外ではシングルタスクと比較して最も推定誤差が小さいサブタスクが異なっていたことから，料理の種類毎に適したサブタスクや適さないサブタスクがあることが示唆された．

3.3.2　複数のサブタスクの利用が推定精度に及ぼす影響

複数のサブタスクを用いたマルチタスクモデルによってシングルタスクより平均して高い推定精度が得られた．これは，サブタスクを複数設定することで様々な情報をバランス良く学習することができたからであると考えられる．しかし，複数のサブタスクを設定した各モデルの推定精度は同程度であり，現状



(a) 入力画像　(b) シングルタスクモデル　(c) マルチタスクモデル

図 7: フィッシュバーガーの画像特徴の可視化結果

Fig. 7　Visualization results of image features for fish burger.

のサブタスクでは更なる魅力度推定精度の向上は困難であると考えられる．そのため，今後はサブタスクの追加や改良等が必要であると考えられる．

3.3.3　料理の種類分類をサブタスクとすることの有効性調査

実験の結果，単一のサブタスクを設定した場合には，料理の種類分類をサブタスクとした場合が最も高精度となった．本サブタスクの有効性をより深く調査するために，NU FOOD 360x10 よりも大規模なデータセットに対して提案手法を適用し，推定精度を調査した．

具体的には，まず，料理画像分類のためのデータセットである UEC FOOD 256 [8] に含まれる 23 種類の料理画像群（図 8）から 2,400 枚を抽出し，各画像に対して魅力度を付与した．ここで，各画像の魅力度は，実験協力者 8〜9 名による 1〜5 の絶対評価の平均値とした．付与された魅力度の例を図 9 に示す．さらに，本実験では作成したデータセットで高精度に魅力度推定するために転移学習をした．具体的には，事前に ImageNet で学習済みの VGG16 を UEC FOOD 256 に含まれる和食 54 種類の画像 7,294 枚すべてを用いてファインチューニングした分類モデルを転移元とし，前述の 23 種類 2,400 枚の魅力度付き画像を用いて魅力度推定モデルを構築した．比較として，シングルタスクモデルとマルチタスクモデルそれぞれについて，ImageNet または ImageNet をファインチューニングした UEC FOOD 256 を転移元とした場合の精度を調査した．本データセットは NU FOOD 360x10 よりも規模が大きいため，Data Augmentation [2] は適用しなかった．各モデルは 100 epoch 学習し，損失関数の重みは表 1 の $N = 1$ のときと同じ値を用いた．その他の実験条件は 3.1 節に記載の条件と同一とした．

実験結果を表 6 に示す．転移元によらず料理の種類分類をサブタスクとするマルチタスク学習によって推定精度が向上することを確認した．結果の詳細な分析の余地はあるものの，料理の種類やデータ数が異なる画像データセットにおいても，料理の種類分類をサブタスクとするマルチタスク学習の有効性が示唆された．

表 6: UEC FOOD 256 [8] に魅力度を付与したデータセットを用いて転移学習した場合の魅力度推定精度（「UEC 和」は，UEC FOOD 256 内の和食画像セットを表す）

Table 6　Accuracy improvement by multi-task learning with transfer learning on an UEC FOOD 256 [8]-based dataset with attractiveness annotations.

	シングルタスク		マルチタスク（単一サブタスク：料理の種類分類）	
転移元	ImageNet	UEC 和	ImageNet	UEC 和
MAE	0.528	0.500	0.448	0.394

4. ま と め

本報告では，料理写真の魅力度推定のためのマルチタスク学習を提案し，その有効性を調査した．実験の結果，料理の種類分類のサブタスクを設定したマルチタスク学習が魅力度推定に有効であることが確認された．今後は，料理毎に適したサブタスクについて分析していく．そのなかで，サブタスクの追加やより大規模な魅力度付きデータセットの利用も検討する．

謝辞　実験用データセット作成にご協力いただいた中京大学高島れんそ氏に感謝する．本研究の一部は，科研費 #20K12038 および MSR Core-12 プログラムによる．

文　献

[1] A. Sato, T. Hirayama, K. Doman, Y. Kawanishi, I. Ide, D. Deguchi, and H. Murase, “Gaze-inspired learning for estimating the attractiveness of a food photo,” Proc. 20th IEEE Int’l Symp. on Multimedia, pp.36–43, Dec. 2018.

[2] T. Hattori, K. Doman, I. Ide, and Y. Mekada, “Application of data augmentation for accurate attractiveness estimation for food photography,” Proc. 11th Workshop on Multimedia for Cooking and Eating Activities, pp.33–40, June 2019.

[3] A.H. Abdulnabi, G. Wang, J. Lu, and K. Jia, “Multi-task CNN model for attribute prediction,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.17, no.11, pp.1949–1959, Nov. 2015.

[4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Proc. 3rd Int’l Conf. on Learning Representations, pp.1–14, May 2015.

[5] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” Proc. 25th Int’l Conf. on Knowledge Discovery and Data Mining, pp.2623–2631, July 2019.

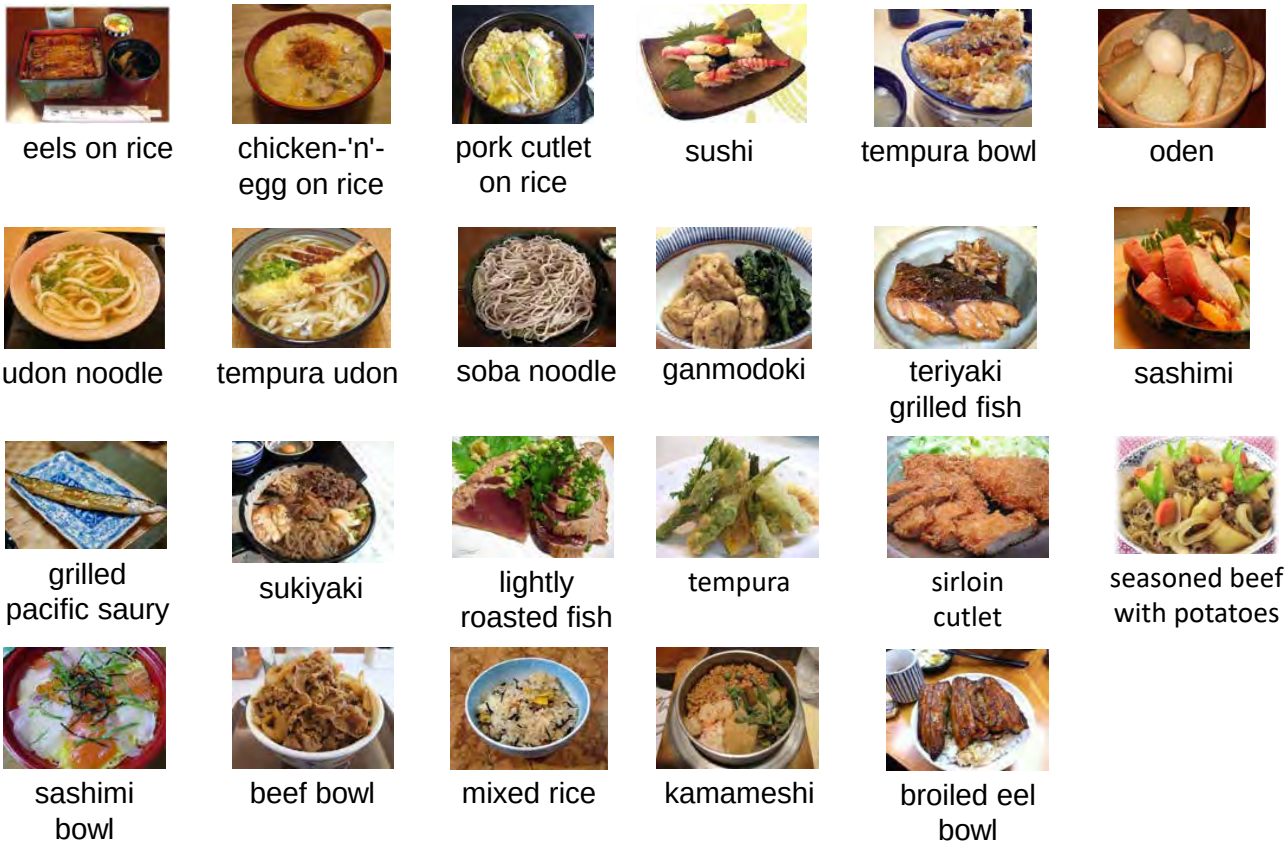


図 8: 魅力度推定のために UEC FOOD 256 データセット [8] から抽出した和食 23 種類

Fig. 8 Twenty three Japanese food categories selected from the UEC FOOD 256 dataset [8] for attractiveness estimation.

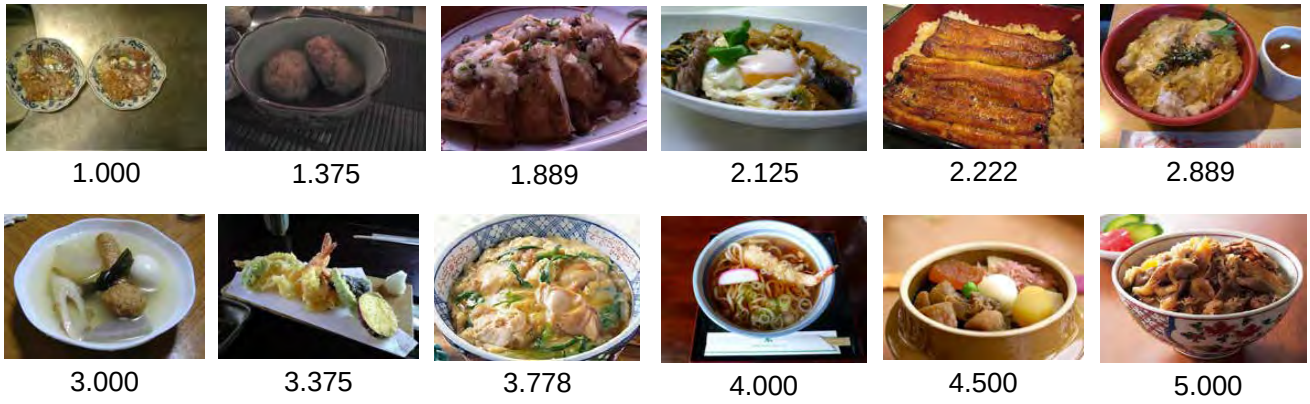


図 9: UEC FOOD 256 データセット [8] から抽出した和食画像に対して付与された魅力度の例（1：低〜5：高）

Fig. 9 Example of attractiveness values for Japanese food images extracted from the UEC FOOD 256 dataset [8] (1: Low to 5: High).

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," Proc. 2009 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.248–255, July 2009.

[7] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Proc. 16th IEEE Int'l Conf. on Computer Vision, pp.618–626, Aug. 2017.

[8] Y. Kawano and K. Yanai, "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation," Proc. ECCV 2014 Workshop on Transferring and Adapting Source Knowledge in Computer Vision, pp.585–597, Apr. 2014.