# 料理写真の高精度な魅力度推定に向けた Data Augmentation

服部　竜実 [†1]　　道満　恵介 [†1]　　井手　一郎 [†2]　　目加田　慶人 [†1]

[†1] 中京大学大学院工学研究科　　[†2] 名古屋大学大学院情報学研究科

## 1　はじめに

SNS や料理レシピサイト等の Web 上に料理写真を投稿する際，料理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 の方がぼかし，撮影角度，被写体の大きさ等の観点から料理が美味しそうに撮影されている．このように，美味しそうな料理写真の撮影は必ずしも容易ではなく，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．そこで本研究では，料理が美味しそうに見える度合いを「魅力度」と定義し，これを料理写真の画像特徴から定量化する技術の開発を目的とする．

これまで我々は，複数の画像特徴を用いた料理写真の魅力度推定手法 [1, 2] を検討してきた．しかし，画像特徴と魅力度の関係をより正確に学習するには，様々な魅力度の画像を含む大規模な画像データセットの作成が必要であった．魅力度付き料理写真データセットを作成する際，結果の再現性や差の弁別性の観点から，魅力度は対比較に基づく手法によって定量化されることが望ましい．しかし，必要となる対比較数は画像枚数に対して指数関数的に増加するため，膨大な時間と労力を要する．

例えば，1,000 枚の画像に対して対比較によって魅力度を付与することを考える．仮に，1 対当たり 5 名の回答を得る場合，${}_{1,000}\mathrm{C}_2 \times 5 = 2,497,500$ 回の対比較が必要となる．ここで，1 名当たり 1,000 回の対比較をしてもらう場合，約 2,500 名の実験参加者が必要となるため，全ての画像に対して対比較によって魅力度を付与することは困難である．更に，画像特徴と魅力度の関係を学習するには画像 1,000 枚では十分とはいえず，対比較のみを用いた大規模な画像データセットの作成は非現実的である．また，絶対評価の場合，個人間および個人内で評価結果にばらつきが生じるため，評価値の信頼性が低くなる．そのため，対比較によって魅力度が付与された少数の画像に対して種々の画像変換を適用することによる Data Augmentation を検討する．ここで，回転や平行移動等の画像変換を単純に適用した場合，元画像と変換後の画像の魅力度は必ずしも一致せず，変換後の画像に対する適切な魅力度の再付与が必要となる．

そこで本研究では，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により，魅力度の再付与を必要としない画像データセットの拡大手法を検討する．本発表では，種々の画像変換が魅力度推定精度に及ぼす影響について分析した結果を報告する．

以降，2 節では，料理写真の魅力度推定手法について概説する．次に，3 節では，種々の画像変換が魅力度推定精度に及ぼす影響を分析した結果について述べる．最後に，4 節で本発表をまとめる．

## 2　料理写真の魅力度推定手法

これまで我々が提案してきた魅力度推定手法 [1] の大まかな処理手順は，図 2 に示すように学習と推定の 2 段階から構成される．学習段階では，予め魅力度が付与された料理写真を基に，回帰の枠組みによって推定器を構築する．回帰モデルには Random Regression Forests [3] を利用し，料理写真の魅力度とその画像特徴量をそれぞれ目的変数，説明変数とする．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写真に対する魅力度を推定する．各段階における画像特徴抽出では，図 3 に示すように，まず入力画像を料理領域 $R_d$ と主食材領域 $R_m$ に分割する．これには，GrabCut [4] の利用およびスマートフォン等の携帯型デバイスを用いたユーザインタラクションを想定する．その後，料理領域 $R_d$ から料理全体の印象に関する画像特徴を抽出し，主食材領域 $R_m$ から主食材の見えに関する画像特徴を抽出する．以降，利用する画像特徴について概説する．

(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

### 2.1　料理全体の印象に関する画像特徴

料理領域 $R_d$ から下記の画像特徴量 $C$，$E$，$A$ をそれぞれ計算する．

#### 2.1.1　色特徴：CIELAB 色空間における色差

まず，料理領域 $R_d$ における最頻出色（$L, a, b$）を計算する．次に，図 4a に示すように料理領域 $R_d$ を放射状に 100 分割し，各部分領域内での最頻出色（$L_i, a_i, b_i$）とその頻度 $F_i$ を計算する．ここで，$i$ は部分領域の番号を表し，$i \in \{1...100\}$ である．その後，特徴量 $C_i$ を下記の式によって計算する．

$$C_i = F_i \sqrt{(L - L_i)^2 + (a - a_i)^2 + (b - b_i)^2} \qquad (1)$$

#### 2.1.2　形状特徴：エッジ強度

まず，図 4b に示すように料理領域 $R_d$ を $10 \times 10$ の格子状に分割し，各部分領域内での最大エッジ強度 $E_j$ を特徴量として利用する．ここで，$j$ は部分領域の番号を表し，$j \in \{1...100\}$ である．

$$E = (E_1, E_2, ..., E_{100}) \qquad (2)$$

#### 2.1.3　色特徴・形状特徴：DeCAF

DeCAF（Deep Convolutional Activation Feature） [5] は ImageNet [6] を用いて学習した畳み込みネットワークにおける全結合層（全 8 中第 7 層目）の出力値である．本手法では，これを色や形状に関する特徴量 $A$ として利用する．

$$A = (A_1, A_2, ..., A_{4096}) \qquad (3)$$

### 2.2　主食材の見えに関する画像特徴

主食材領域 $R_m$ から下記の画像特徴量 $S$，$P_x$，$P_y$，$O$，$M$ をそれぞれ計算する．

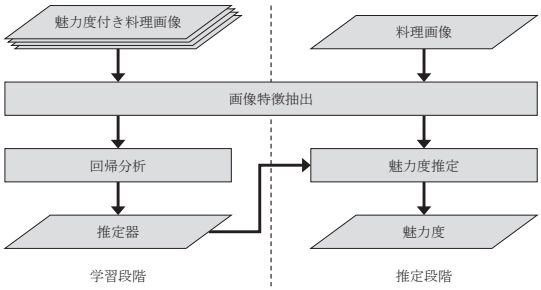



図 2: 魅力度推定の処理手順

(a) 入力画像　(b) 料理領域　(c) 主食材領域

図 3: 領域分割結果の例

#### 2.2.1　サイズ特徴：主食材の大きさ

料理領域 $R_d$ に対する主食材領域 $R_m$ の面積比 $S$ を特徴量として利用する．

$$S = \frac{|R_m|}{|R_d|} \tag{4}$$

#### 2.2.2　位置特徴：主食材の相対位置

料理領域 $R_d$ の重心（$x_d, y_d$）と主食材領域 $R_m$ の重心（$x_m, y_m$）の水平・垂直方向の差 $P_x$，$P_y$ を特徴量として利用する．

$$P_x = x_d - x_m \tag{5}$$

$$P_y = y_d - y_m \tag{6}$$

#### 2.2.3　形状特徴：主食材の方向ヒストグラムとモーメント

主食材領域 $R_m$ における勾配方向を 36 分割したヒストグラムを特徴量 $O$ として利用する．

$$O = (O_1, O_2, ..., O_{36}) \tag{7}$$

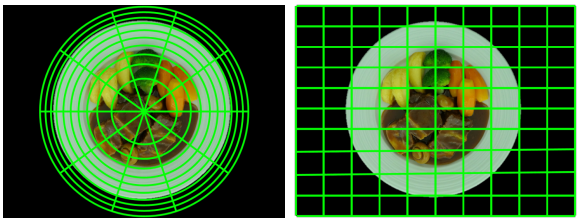
また，$O$ の第 1 次〜第 4 次モーメントを特徴量 $M$ として利用する．

$$M = (M_1, M_2, M_3, M_4) \tag{8}$$

ここで，$M_1$，$M_2$，$M_3$，$M_4$ はそれぞれ $O$ の平均，分散，歪度，尖度である．

## 3　評価実験

Data Augmentation の有効性を検証するため，種々の画像変換が魅力度推定精度に及ぼす影響を定量的に分析した．具体的には，予め魅力度が付与された画像データセットに対して種々の画像変換を適用し，画像特徴と推定精度の関係を分析した．また，各画像変換を単体で適用した場合と同時に適用した場合の精度を比較した．以降，実験の方法および結果を述べ，考察する．



(a) 色特徴　(b) 形状特徴

図 4: 特徴量の抽出領域

### 3.1　方法

本実験では，魅力度付き料理写真データセット NU FOOD 360x10 [1]を利用した．このデータセットは，10 種類の料理を対象とし，各料理について 3 種類の仰角（$30°, 60°, 90°$）と 12 種類の回転角（$0°, 30°, \ldots, 330°$）から料理サンプルを撮影した計 36 枚の画像からなる．また，各料理について，Thurstone の一対比較法 [7] によって魅力度の目標値（0〜1 の連続値）が料理写真毎に付与されている．このデータセットに対して，まず，下記の画像変換をそれぞれ適用した．各画像変換によって生成された画像の例を図 5 に示す．

- 回転：画像を時計回りまたは反時計回りに $\theta°$ 回転（$0.1 \leq \theta \leq 5.0$）
- 拡縮:画像を縦横 $s$ 倍に拡縮（$0.95 \leq s \leq 1.05$ かつ $s \neq 1.00$）
- 平行移動：画像を横に $d_x$ 画素かつ縦に $d_y$ 画素移動（$0 \leq |d_x| \leq 5$ かつ $0 \leq |d_y| \leq 5$ かつ $|d_x| + |d_y| \neq 0$）
- ランダムノイズ：RGB のチャネル毎に画像の各画素値に $n$ を付与（$0 \leq |n| \leq 5$）

次に，オリジナル画像セット $D_{\text{orig}}$ に対して 4 種類の画像変換をそれぞれ単独に適用した画像セット $D_{\text{rotate}}$，$D_{\text{scale}}$，$D_{\text{shift}}$，$D_{\text{noise}}$ と，これら全てを組み合わせた画像セット $D_{\text{all}}$，4 種類の画像変換を同時に適用した画像セット $D_{\text{comb}}$ を作成した．本実験で作成した各画像セットに含まれる画像枚数を表 1 に示す．そして，各画像セットに対して 2 節で述べた手法を用いた leave-one-out を適用し，魅力度の推定値と目標値の間の平均絶対誤差（MAE）を評価した．

### 3.2　結果

各料理における画像セット毎の結果を表 2 に示す．オリジナル画像セット $D_{\text{orig}}$ に対する MAE の平均値は 0.088 であった．各画像変換を同時適用した画像セット $D_{\text{comb}}$ に対する MAE の平均値は 0.079 であった．各画像変換をそれぞれ単独に適用した画像セットを組み合わせた画像セット $D_{\text{all}}$ に対する MAE の平均値は 0.077 であった．これにより，種々の画像変換の統合利用が高精度な魅力度推定に有効であることが確認された．また，$D_{\text{rotate}}$，$D_{\text{scale}}$，$D_{\text{shift}}$，$D_{\text{noise}}$ に対する MAE の平均値はそれぞれ 0.083，0.084，0.087，0.090 であった．

### 3.3　考察

各画像変換をそれぞれ単独に適用した画像セット毎の結果について，$D_{\text{rotate}}$，$D_{\text{scale}}$，$D_{\text{shift}}$ では平均して $D_{\text{orig}}$ よりも小さい推定誤差を示した．魅力度に影響しない範囲で Data Augmentation

[1]NU FOOD 360x10: http://www.murase.is.i.nagoya-u.ac.jp/nufood/

(a) オリジナル

(b) 回転

(c) 拡縮

(d) 平行移動

(e) ランダムノイズ

図 5: 各画像変換によって生成された画像の例

Table 1: 各画像セットに含まれる画像枚数 [枚]

| $D_{\mathrm{orig}}$ | $D_{\mathrm{rotate}}$ | $D_{\mathrm{scale}}$ | $D_{\mathrm{shift}}$ | $D_{\mathrm{noise}}$ | $D_{\mathrm{comb}}$ | $D_{\mathrm{all}}$ |
|---|---|---|---|---|---|---|
| 36 | 396 | 396 | 396 | 396 | 1,476 | 1,476 |

したことにより，各画像変換に対して頑健な推定器を構築できたと考えられる．これに対して，$D_{\mathrm{noise}}$ では平均して $D_{\mathrm{orig}}$ よりも大きい推定誤差を示した．これは，ノイズ量が大きすぎたことにより，本来の魅力度と異なる分布の画像を生成・学習してしまったためと考えられる．

また，画像変換を統合利用した画像セットについて，$D_{\mathrm{all}}$ では平均して $D_{\mathrm{comb}}$ よりも小さい推定誤差を示した．これは，ほとんどの画像変換はそれぞれ単独では魅力度に影響しない範囲で適用されていたが，同時適用したことで総変化量が大きくなってしまったためと考えられる．図 6 は元画像に対して画像変換を単独に適用した場合と同時に適用した場合の比較であるが，図 6b のように単独に適用した場合は，画像の変化が小さいため，元画像と変換後の画像の見えの違いを知覚できない．一方，図 6c のように同時に適用した場合は，画像が大きく変化しており，見えの違いを知覚できるため，元画像と変換後の画像の魅力度は一致しないと考えられる．

これらの結果から，より高品質な Data Augmentation の実現には，同時適用時の総変化量が大きくなりすぎないようにする方法の検討が必要であること，および，画像変換毎の最適な変化量の網羅的な調査が必要であることが示唆された．

## 4 おわりに

本発表では，魅力度の再付与を必要としない画像データセットの拡大を目的とし，種々の画像変換が魅力度推定精度に及ぼす影響について定量的に分析した結果を報告した．予め魅力度が付与された画像データセットに対して種々の画像変換を適用し，画像特徴と推定精度の関係を分析した．また，各画像変換を単体で適用した場合と同時に適用した場合の精度を比較した．実験の結果，画像変換による Data Augmentation の有効性が確認されたが，より高品質な Data Augmentation の実現には，同時適用時の総変化量が大きくなりすぎないようにする方法の検討が必要であること，および，画像変換毎の最適な変化量の網羅的な調査が必要であることが示唆された．

今後は，各画像変換における料理毎の最適な変化量について詳しく分析していく．また，より複雑な画像変換方法として，Generative Adversarial Nerworks [8] や Adversarial Examples [9] 等の利用も検討していく．

**謝辞** 本研究の一部は JSPS 科研費 JP17K12719 および 16H02846，Microsoft CORE-12 プログラムによる．

## 参考文献

[1] K. Takahashi, K. Doman, T. Hirayama, Y. Kawanishi, I. Ide, D. Deguchi and H. Murase: Estimation of the attractiveness of food photography focusing on main ingredients, Proc. 9th Workshop on Multimedia for Cooking and Eating Activities (CEA), pp.1–6 (2017).

[2] T. Hattori, K. Doman, I. Ide and Y. Mekada: A study on the factors affecting the attractiveness of food photography, Proc. 10th Workshop on Multimedia for Cooking and Eating Activities (CEA), pp.25–28 (2018).

[3] A. Liaw and M. Wiener: Classification and regression by randomForest, R News, vol.2, no.3, pp.18–22 (2002).

Table 2: 各料理における画像セット毎の推定精度（MAE：平均絶対誤差）

| 料理 | $D_{\mathrm{orig}}$ | $D_{\mathrm{rotate}}$ | $D_{\mathrm{scale}}$ | $D_{\mathrm{shift}}$ | $D_{\mathrm{noise}}$ | $D_{\mathrm{comb}}$ | $D_{\mathrm{all}}$ |
|---|---|---|---|---|---|---|---|
| 鰹のたたき | 0.133 | 0.134 | 0.122 | 0.111 | 0.157 | 0.120 | **0.107** |
| カレーライス | 0.086 | 0.079 | **0.077** | 0.111 | 0.085 | 0.086 | 0.078 |
| 鰻丼 | 0.068 | 0.063 | **0.062** | 0.069 | 0.105 | 0.063 | 0.066 |
| ビーフシチュー | 0.097 | 0.075 | 0.090 | 0.072 | **0.063** | 0.075 | 0.077 |
| ハンバーグ | **0.094** | 0.109 | 0.121 | 0.121 | 0.110 | 0.102 | 0.096 |
| 天丼 | 0.115 | 0.109 | 0.116 | 0.118 | 0.119 | **0.102** | 0.114 |
| カツ丼 | 0.097 | 0.098 | 0.099 | 0.097 | 0.095 | **0.084** | 0.093 |
| 鉄火丼 | 0.048 | 0.038 | **0.034** | 0.035 | 0.046 | 0.041 | 0.036 |
| チーズバーガー | 0.061 | 0.045 | 0.047 | 0.071 | 0.050 | 0.045 | **0.043** |
| フィッシュバーガー | 0.088 | 0.086 | 0.071 | 0.067 | 0.071 | 0.068 | **0.063** |
| 平均 | 0.088 | 0.083 | 0.084 | 0.087 | 0.090 | 0.079 | **0.077** |

(a) オリジナル　(b) 単独に適用した場合（平行移動）　(c) 同時に適用した場合

図 6: 画像変換の同時適用によって元画像が大きく変化してしまう例


[4] C. Rother, V. Kolmogorov and A. Blake: GrabCut — Interactive foreground extraction using iterated graph cuts, ACM Trans. on Graphics —Proc. ACM SIGGRAPH 2004, vol.23, no.3, pp.309–314 (2004).

[5] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng and T. Darrell: DeCAF: A deep convolutional activation feature for generic visual recognition, Proc. 31st International Conference on Machine Learning, pp.647–655 (2014).

[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 Conference on Computer Vision and Pattern Recognition, pp.248–255 (2009).

[7] L. L. Thurstone: Psychophysical analysis, American Journal of Psychology, vol.38, no.3, pp.368–389 (1927).

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio: Generative Adversarial Nets, Advances in Neural Information Processing Systems 27, pp.2672–2680 (2014).

[9] T. Miyato, S. Maeda, M. Koyama, K. Nakae and S. Ishii: Distributional smoothing with virtual adversarial training, Computing Research Repository, arXiv:1507.00677 (2015).