料理写真の魅力度推定精度向上のための視線停留分布に対するデータ拡張手法の検討

○宮崎光明[1)]，服部竜実[1)]，道満恵介[1)]，
平山高嗣[2)]，川西康友[3)]，井手一郎[4),3)]，目加田慶人[1)]

○ Mitsuaki MIYAZAKI [1)], Tatsumi HATTORI [1)], Keisuke DOMAN [1)],
Takatsugu HIRAYAMA [2)], Yasutomo KAWANISHI [3)], Ichiro IDE [4),3)],
and Yoshito MEKADA [1)]

1) 中京大学 大学院工学研究科
2) 名古屋大学 未来社会創造機構
3) 名古屋大学 大学院情報学研究科
4) 名古屋大学 数理・データ科学教育研究センター

{miyazaki.m, hattori.t}@md.sist.chukyo-u.ac.jp, {kdoman, y-mekada}@sist.chukyo-u.ac.jp,
takatsugu.hirayama@nagoya-u.jp, {kawanishi, ide}@i.nagoya-u.ac.jp

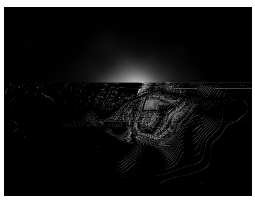
＜要約＞本発表では，料理写真の魅力度推定精度の向上に向けた視線情報付き画像データセットの拡張について検討した結果を報告する．これまで我々は，料理写真を評価する際の視線情報を考慮した魅力度推定手法を提案した．しかし，視線情報付き画像データセットの規模が小さく，視線情報利用の有効性を検証するには不十分であった．そこで本研究では，従来研究の料理画像に対する Data Augmentation 手法の有効性から着想を得て，視線計測実験によって得られた視線停留分布に対して Data Augmentation を適用する．これにより，追加の視線計測実験を行わずに大規模な視線情報付き画像データセットを作成する手法を検討する．実験によりその有効性を確認した．
＜キーワード＞料理写真，魅力度，視線，回帰分析，データ拡張

1　はじめに

SNS や料理レシピサイト等に投稿する際の料理写真は，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 1(a) よりも，図 1(b) の方が撮影角度，ぼかし，被写体の大きさ等の観点から美味しそうに撮影されている．しかし，料理写真を美味しそうに撮影すること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．そこで我々は，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，これを画像特徴から高精度に推定する技術の開発を目的として研究している．そのなかでこれまで我々が行った研究を二つ示す．まず，佐藤らは CNN から抽出された特徴マップの重みとして視線停留分布を利用する手法 [1] を提案した．視線停留分布の例を図 2 に示す．視線停留分布とは，視線の累積停留時間の空間分布のことであり，料理写真を評価する際に人が空間中のどの

(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



(a) 料理画像　(b) 視線停留分布

図 2　料理画像に対する視線停留分布の例

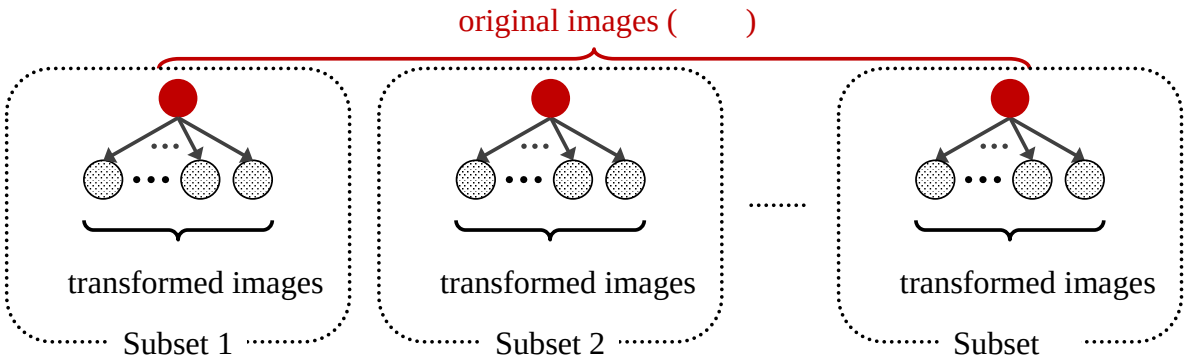


図 6　Leave-one-subset-out 法の概要

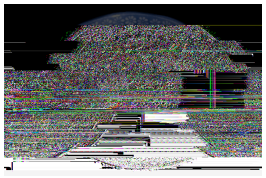
表 1　データ拡張の有無による推定精度（MAE：平均絶対誤差）

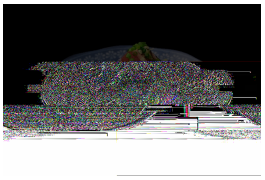
料理	データセット	
	G_{orig}（ベースライン手法）	G_{DA}（提案手法）
鰹のたたき	0.138	0.154
カレーライス	0.113	0.094
鰻丼	0.101	0.066
ビーフシチュー	0.106	0.114
ハンバーグ	0.129	0.120
天丼	0.101	0.092
カツ丼	0.145	0.129
鉄火丼	0.095	0.052
チーズバーガー	0.100	0.100
フィッシュバーガー	0.112	0.087
平均	0.114	0.101

0.7 以上の画像については平均して推定誤差が小さくなり，それ以外の画像については推定誤差が大きくなった．鰹のたたきにおける魅力度が高い画像と低い画像の視線停留分布を重畳表示した例を図 7 に示す．魅力度が高い画像では鰹と僅かに生姜の領域に視線が停留しており，魅力度が低い画像では鰹の領域のみに視線が停留していることが分かる．このことから，魅力度が高い画像とそれ以外の画像では微小な画像変換によってそれぞれ，鰹と生姜を含んだ領域と鰹の領域の視線停留分布が拡充されたことになる．よって，魅力度が高い画像ではデータ拡張によってベースライン手法より鰹と生姜の位置関係を学習することができ，推定誤差が小さくなったと考えられる．このことから，鰹のたたきの魅力度推定では食材の配置や料理の構成を考慮した上で視線停留分布をデータ拡張する手法が有効であると考えられる．

4.3.2　Data Augmentation の適用対象の変更が推定精度に及ぼす影響

従来のデータ拡張 [2] と比較するために，Data Augmentation の適用対象を変更した．料理画像に対して Data Augmentation を適用したデータセット I_{DA} と視線停留分布と料理画像へ同時に Data Augmentation を適用したデータセット $(G+I)_{\mathrm{DA}}$ を作成した．なお，$(G+I)_{\mathrm{DA}}$ では視線停留分布と料理画像を一対一に対応させて画像変換を行った．これらを用いて魅力度推定した結果を表 2 に示す．I_{DA} では MAE の平均値は 0.087 であり，$(G+I)_{\mathrm{DA}}$ では MAE の平均値は 0.089 であった．$(G+I)_{\mathrm{DA}}$ は G_{DA} よりも推定誤差が小さくなったことから，視線停留分布と料理画像に対する Data Augmentation の同時適用の有効性が確認された．しかし，G_{DA} と $(G+I)_{\mathrm{DA}}$ は I_{DA} よりも推定誤差が大きくなった．このことから，視線停留分布に適したデー


(a) 魅力度：1.00


(b) 魅力度：0.83

(c) 魅力度：0.10

(d) 魅力度：0.00

図 7　鰹のたたきの魅力度の例

表 2　Data Augmentation の適用対象の変更による推定精度

料理	データセット		
	G_{DA}	I_{DA}	$(G+I)_{\mathrm{DA}}$
鰹のたたき	0.154	0.116	0.134
カレーライス	0.094	0.078	0.073
鰻丼	0.066	0.076	0.081
ビーフシチュー	0.114	0.076	0.077
ハンバーグ	0.120	0.134	0.129
天丼	0.092	0.097	0.083
カツ丼	0.129	0.103	0.109
鉄火丼	0.052	0.048	0.048
チーズバーガー	0.100	0.072	0.078
フィッシュバーガー	0.087	0.073	0.073
平均	0.101	0.087	0.089

表 3　画像変換パラメータの値の変更による推定精度

料理	データセット	
	G_{DA}	G'_{DA}
鰹のたたき	0.154	0.154
カレーライス	0.094	0.089
鰻丼	0.066	0.063
ビーフシチュー	0.114	0.108
ハンバーグ	0.120	0.117
天丼	0.092	0.087
カツ丼	0.129	0.129
鉄火丼	0.052	0.052
チーズバーガー	0.100	0.099
フィッシュバーガー	0.087	0.089
平均	0.101	0.099

タ拡張方法は，料理画像と異なることが示唆された．

4.3.3　画像変換パラメータの値が推定精度に及ぼす影響

4.3.2 項の結果から，視線停留分布に適したデータ拡張方法は，料理画像と異なることが示唆された．そこで，本項と 4.3.4 項では，画像変換方法を変更した結果を報告する．まず本項では，画像変換パラメータの値を変更した．従来手法 [2] では，料理画像に対して画像変換を適用したときに，元画像と見え方の変化を知覚できない範囲でデータ拡張を行ったが，視線停留分布は選好者の注視領域や注視方法等の観点から様々であるため，画像変換を適用したときに元画像の見え方と一致する必要はない．そのため，従来よりも広い範囲で画像変換を適用しても実測データとしてみなすことが可能であると考えられる．そこで，上記の画像変換パラメータの値全てを 1.50 倍したデータセット G'_{DA} を作成した．これらを用いて魅力度推定した結果を表 3 に示す．G'_{DA} では MAE の平均値は 0.099 であった．G_{DA} は料理写真のデータ拡張における最適な画像変換パラメータの値を用いているが，G'_{DA} は G_{DA} よりも推定誤差が小さくなった点から，視線計測の実測データとしてみなせる範囲は，料理写真の見え方の変化を許容しない範囲よりも広い特徴空間の範囲にあることが示唆された．

4.3.4　ランダムノイズの除去が推定精度に及ぼす影響

視線停留分布に適用する画像変換を変更した．本実験では計測誤差や選好者間の視線の差異として許容できる範囲でデータ拡張を行ったが，それらの観点ではランダムノイズは現れないと考えられる．そのため，種々の画像変換からランダムノイズを除外したデータセット $G_{\mathrm{DA(w/o\ noise)}}$ と $G_{\mathrm{DA(w/o\ noise)}}$ の画像変換パラメータを 1.50 倍したデータセット $G'_{\mathrm{DA(w/o\ noise)}}$ を作成した．これらを用いて魅力度推定した結果を表 4 に示す．$G_{\mathrm{DA(w/o\ noise)}}$ の MAE の平均値は 0.097 であり，$G'_{\mathrm{DA(w/o\ noise)}}$ の MAE の平均値は 0.098 であった．この結果から，視線停留分布の画像変換において適切な画像変換は料理画像と異なることを確認し，ランダム

表 4　ランダムノイズの有無による推定精度

料理	データセット		
	G_{DA}	$G_{\mathrm{DA(w/o\ noise)}}$	$G'_{\mathrm{DA(w/o\ noise)}}$
鰹のたたき	0.154	0.143	0.145
カレーライス	0.094	0.085	0.085
鰻丼	0.066	0.073	0.074
ビーフシチュー	0.114	0.101	0.103
ハンバーグ	0.120	0.113	0.117
天丼	0.092	0.091	0.088
カツ丼	0.129	0.135	0.137
鉄火丼	0.052	0.046	0.045
チーズバーガー	0.100	0.097	0.099
フィッシュバーガー	0.087	0.080	0.083
平均	0.101	0.097	0.098

表 5　追加で作成したデータセットに含まれるデータ拡張方法と画像変換パラメータ

データセット	Data Augmentation の適用対象		ランダムノイズ付与		画像変換パラメータ [倍]
	視線停留分布	料理画像	あり	なし	
I_{DA}		✓	✓		1.00
$(G+I)_{\mathrm{DA}}$	✓	✓	✓		1.00
G'_{DA}	✓		✓		1.50
$G_{\mathrm{DA(w/o\ noise)}}$	✓			✓	1.00
$G'_{\mathrm{DA(w/o\ noise)}}$	✓			✓	1.50

ノイズの付与は不要であることが示唆された．

4.3.2 項以降で新たに作成した各データセットの概要を表 5 にまとめる．これらのデータセットを用いて魅力度推定を行った結果，全てのデータセットで MAE の平均が G_{DA} の MAE の平均より小さい値となり，推定精度の向上を確認した．これらの結果から，視線停留分布における最適なデータ拡張方法は従来のデータ拡張 [2] と異なることを確認した．

5　まとめ

本発表では，視線停留分布のデータ拡張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した．実験の結果，データ拡張が料理写真の高精度な魅力度推定に有効であることが確認された．しかし，追加実験から従来のデータ拡張手法 [2] は，視線停留分布において最適ではないことが示された．今後は，視線停留分布の画像変換における料理毎の最適な変換方法について詳しく分析していく．そのなかで，人間の眼球運動の性質など，視線の特性を考慮したデータ拡張も検討していく．

謝辞　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．

参考文献

[1] Akinori Sato, Takatsugu Hirayama, Keisuke Doman, Yasutomo Kawanishi, Ichiro Ide, and Hiroshi Murase: “Gaze-inspired learning for estimating the attractiveness of a food photo”, Proceedings of 20th IEEE International Symposium on Multimedia, pp.36–43, Dec. 2018

[2] Tatsumi Hattori, Keisuke Doman, Ichiro Ide, and Yoshito Mekada: “Application of data augmentation for accurate attractiveness estimation for food photography”, Proceedings of 11th Workshop on Multimedia for Cooking and Eating Activities, pp.33–40, June 2019

[3] Louis Leon Thurstone: “Psychophysical analysis”, American Journal of Psychology, vol.38, no.3, pp.368–389, 1927