解説 認識・検出

スタイル特徴を利用したDNNによる印象推定に寄与する画像領域の可視化

Visualization of Image Region Contributing to Impression Estimation by DNN Using Style Features

関西学院大学
飛谷 謙介・谷　伊織・橋本　翔・長田 典子

はじめに

プロダクトデザインにおいて、ユーザのニーズを的確に把握することは重要である。近年、ユーザは使い心地や快適性などの感性的なニーズといった機能性だけでない付加価値も重要なものと考えられている[1]。このような感性的なニーズを把握するための、信頼性が高く、有効な方法論の一つとして感性工学による手法があげられ、さまざまなドメインに対して適用されている[2) 3)]。当該手法は、心理学的実験と統計解析を活用することで、感性情報の特徴である個人のばらつき、状況依存性、並びに潜在性（因果関係の希薄さ）といった諸課題[4]を解決し、感性的なニーズを感性指標として顕在化させることを可能にする[5]。さらに、構築した感性指標と対象とするプロダクトの物理特性との関係をモデル化することで、感性的なニーズにつながる「印象」や「価値」に対する物理的な要因の推定が可能になる。これにより、プロダクトデザインの現場において、効率的な上流工程へのフィードバックが実現する。しかしながら、高精度な感性指標を構築するためには心理実験による主観データの取得とその分析が必要であるため、人的および時間的な負荷が高いという問題がある。このような問題に対して機械学習を活用する研究が進められている[13]。

感性情報を含むデータセット[6) 7) 8)]を深層学習などの機械学習手法で扱う場合、個人によって評価がばらつくといった感性情報固有の特徴を考慮した学習手法が必要である。さらに、プロダクトデザインへのフィードバックを目的とするならば、学習における判断根拠を可視化する必要がある。深層学習は一般物体認識において人の認識精度を超える高い性能を達成しているが[9]、深層学習がどのような観点で認識しているのかは理論的に解明されていなかったが、近年、その判断根拠を明確にするための研究が盛んに行われてきている[10) 11)]。

本稿では、感性情報を含むデータに機械学習手法を活用した研究事例として、プロダクトの画像の印象評価において生じる「個人のばらつき」を包含した高精度な印象推定を深層学習モデルを用いて実現する手法を紹介したのち、構築したモデルにおいてプロダクトの印象に寄与する画像領域の可視化を行うことで、その妥当性を確認する。

スタイル特徴を学習するCNNの構築

本章では、高精度な感性指標を構築するために用いる印象推定モデルの構築手法について述べる。まず、構築する印象推定モデルについての概説し、その後、具体的な構築手法を述べる。

プロダクト画像に対する印象推定モデル

本研究ではCNNを用いて印象推定を実現する。CNNへの入力データとしては学習済みVGG-19を通して抽出したスタイル特徴を利用する。これまでに著者らは衣服の柄を対象とした印象推定モデルを構築した際にスタイル特徴を使用し、印象推

定に対する有効性を確認している[12]。スタイル特徴を入力データとして使用する場合、深いPooling層から抽出されたスタイル特徴は非常に高次元であるため、データセットのサンプル数が少ない場合、学習の際に過学習することが予想される。そこで本手法では、浅いpooling層から抽出される低次元のスタイル特徴を用い、プロダクト画像から喚起される印象の推定を実現する。

対象データセット

本手法では、定量化された印象（評価値）が付与された腕時計の画像データセット[13]を学習に用いる。腕時計はデザインを構成する要素が少ないため、デザイン要素と印象との関係が解釈可能で、提案手法の有用性を示しやすいと考えられる。

また、本研究では視覚的な印象を扱うため、データセットに付与されている印象のうち、「上品だ、可愛らしい、小さい」を対象とした。

本データセットは、画像1枚あたり20人分の評価値が付与された計2000枚の画像で構成されている。全画像において背景が白色であり、腕時計が中央に位置している。また、評価値は、主観評価実験を通して1から7の7段階で付与されている。実験は、クラウドソーシングサービスを利用して実施された。実験では、各実験参加者は画像を1枚ずつ観察し、評価項目に対して「非常に当てはまらない」、「当てはまらない」、「やや当てはまらない」、「どちらでもない」、「やや当てはまる」、「当てはまる」、「非常に当てはまる」の7段階で評価した。その際、画像の呈示順序は実験参加者ごとにランダムとした。本研究では、20人分の評価結果から、7段階の評価値を1から7の値に変換し、縦軸と横軸にそれぞれ人数（0－1に正規化）と評価値（1－7）をとった離散確率分布の形式で扱う。データセットの代表例を第1図に示す。

スタイル特徴

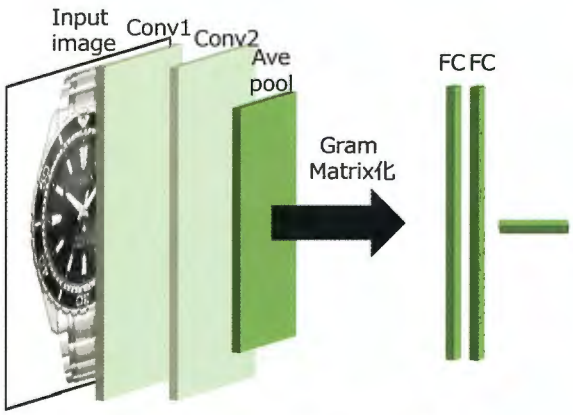
スタイル特徴は、一般物体認識に用いられる畳み込みニューラルネットワークであるVGG-19の中間層から出力される特徴マップの相互相関行列であり、画像中の色情報やパターン情報などの詳細な見た目を表現する特徴量と言われている[14]。特

第1図　データセットの代表例

徴次元数は、VGG-19のpooling層１、２、３、４で抽出されることから、それぞれ64×64、128×128、256×256、512×512である。

CNNの構築

本研究で扱うCNNは、プロダクト画像を入力データとし、スタイル特徴を抽出、全結合層を挟んで出力層でクラス分類を行う。その際、クラス分類におけるクラス番号は評価値と対応している。構築したCNNの構造を第２図、第１表にそれぞれ示す。



第２図　構築したCNNの概念図

第１表　構築したCNNの構造

Layer	Input size	Output size	Kernel	Stride
Conv1	224 × 224 × 3	224 × 224 × 64	3 × 3	1
Conv2	224 × 224 × 64	224 × 224 × 64	3 × 3	1
Ave pool	224 × 224 × 64	112 × 112 × 64	2 × 2	2
FC1	1 × 1 × 4096	1 × 1 × 4096		
FC2	1 × 1 × 4096	1 × 1 × 4096		
FC3	1 × 1 × 4096	1 × 1 × 7		

CNNの学習における最適化手法はadam[15]を使用し、バッチサイズは128とする。全結合層での重みの初期値にはHeの初期値[16]を適用し、学習率の初期値は0.000001とする。活性化関数にはRectified Linear Unit（ReLU）関数[17]を使用し、全結合層ではBatch Normalization[18]を適用する。また、出力層ではsoftmax関数を使用し、クラス番号iに対する尤度P_iを算出する。さらに、算出した尤度P_iと画像の教師ラベルから損失関数による誤差を算出し、これを最小化するようにCNNの重みを更新することで学習を行う。

損失関数の選定

深層学習において、パラメータを最適化する際のアルゴリズムだけでなく、損失関数の選定も推定精度に大きく関わってくる。本手法では画像から喚起される印象を離散確率分布として扱うため、当該形式に適した損失関数を選ぶ必要がある。

そこで、クラス分類問題でよく用いられるクロスエントロピー誤差（CE）、およびJinらにより提案された、確率分布の差を表すJensen-Shannon情報量に基づく損失関数（CJS）[19]とで精度比較を行う。CEおよびCJSをそれぞれ式(1)、式(2)に示す。誤差の算出方法は評価値の分布と推定値の分布とのMSEとする。

$$E = -\sum_{i=1}^{Z}[y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)] \quad \cdots(1)$$

$$E = \frac{1}{2}\left[\sum_{i=1}^{Z} Y_1(i)\ln\left(\frac{Y_1(i)}{Y^s}\right) + \sum_{i=1}^{Z} Y_2(i)\ln\left(\frac{Y_2(i)}{Y^s}\right)\right]$$

$$Y^s = \frac{1}{2}Y_1(i) + \frac{1}{2}Y_2(i)$$

$$Y(i) = \sum_{j=1}^{i} y(j), \quad i = 1, \ldots, Z \quad \cdots(2)$$

ここでY_1およびY_2はそれぞれ推定値の分布と印象値の分布、Zは評価の段階数であり、$\hat{y}_i$およびy_iはそれぞれ推定値の分布、評価値の分布におけるクラス番号iに対応した値である。

比較に用いるDNNの詳細を第２表に示す。構築したDNNはスタイル特徴を入力データとし、隠れ層は全結合層、出力層はsoftmax関数によるクラス分類とする。また、DNNの学習方法は2.4節で述べたものと同様である。

第２表　損失関数比較に用いるDNN

Layer	Input size	Output size
FC1	1 × 1 × 4096	1 × 1 × 4096
FC2	1 × 1 × 4096	1 × 1 × 4096
FC3	1 × 1 × 4096	1 × 1 × 7

得られた推定誤差を第３表に示す。クロスエントロピー誤差を用いたモデルの推定誤差の方が低いことがわかる。したがって提案手法では、損失関数としてクロスエントロピー誤差を用いる。

第３表　損失関数比較における推定誤差

	CE	CJS
MSE	0.117	0.136

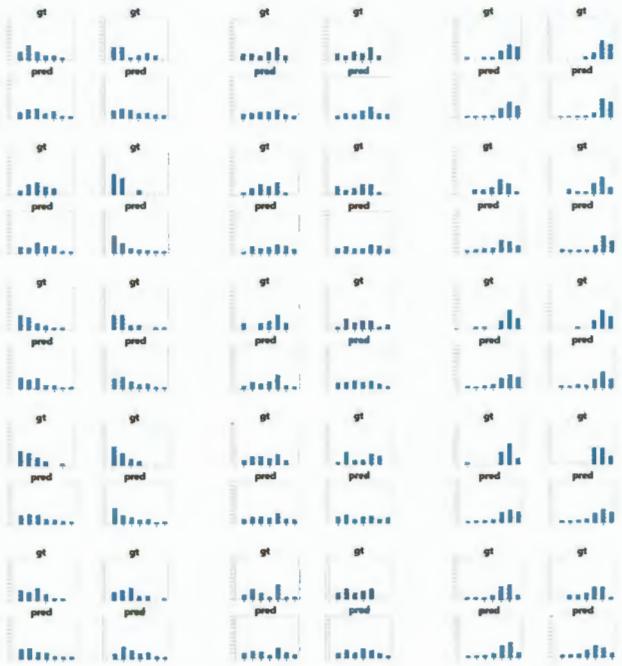
印象推定結果

提案手法と既存手法との印象推定における精度比較を行った。比較には、10-fold交差検証における推定誤差の平均を用いた。推定誤差の算出には推定された分布と評価値の分布とのMSE（Mean Squared Error）を用いる。

比較対象はVGG19（ImageNet[20]にて学習済み）のfine tuningにより全結合層の重みを再学習したものとした（VGG19）。また、fine tuningにおける学習方法は2.4節で述べたものと同様とする。得られた推定誤差の平均を第４表に示す。第４表から提案手法の方が、推定誤差が小さいことが確認され、提案手法がプロダクトの画像の印象を推定するうえで有効であることが示唆された。さらに、推定結果が印象の強弱に対応したクラスの離散確率分布で出力されることから「個人のばらつき」を包含した印象推定モデルを構築できたと言える。印象推定結果の代表例を第３図に示す。なお、図中gtは評価値の分布、predは推定値の分布である。

第４表　印象推定における推定誤差の平均

	本手法	VGG19
MSE	0.094	0.192



第３図　印象推定結果の代表例

印象に寄与する画像領域の可視化

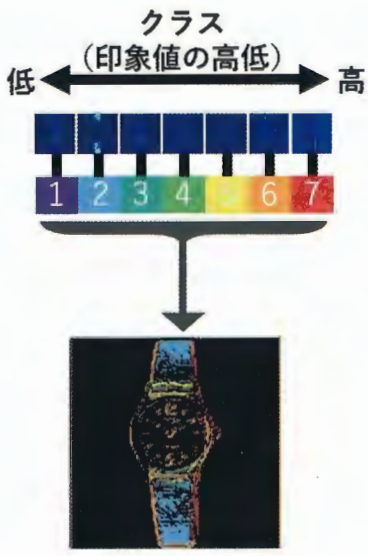
本章では、前章で構築したCNNにおけるプロダクトの印象推定に寄与する画像領域の可視化手法、およびその妥当性について述べる。まず、可視化の流れについて述べ、その際に使用する手法について説明し、その後、得られた可視化結果に

対する考察を加える。

可視化フロー

可視化フローを第４図に示す。本手法は以下の二つの処理で構成されている。まず、(i)各印象値に寄与する画像領域の可視化を行う。その後、(ii)寄与度の高い画像領域の統合を行う。

(i)ではクラス分類における各クラス（印象の強弱）に寄与する画像領域をGrad-CAM[11]により可視化を行い、クラス数分の印象の強度に対応したヒートマップ画像を取得する。(ii)では(i)で得られたクラス数分のヒートマップ画像を用いて画素毎にその強度比較を行い、最も寄与度の高いクラスに対応する色（既定）をその画素に付加する。以上の処理を行い得られた画像を「ヒートマップ統合画像」とする。また、一定の閾値以下の画素は黒色とする。



第４図　可視化フロー

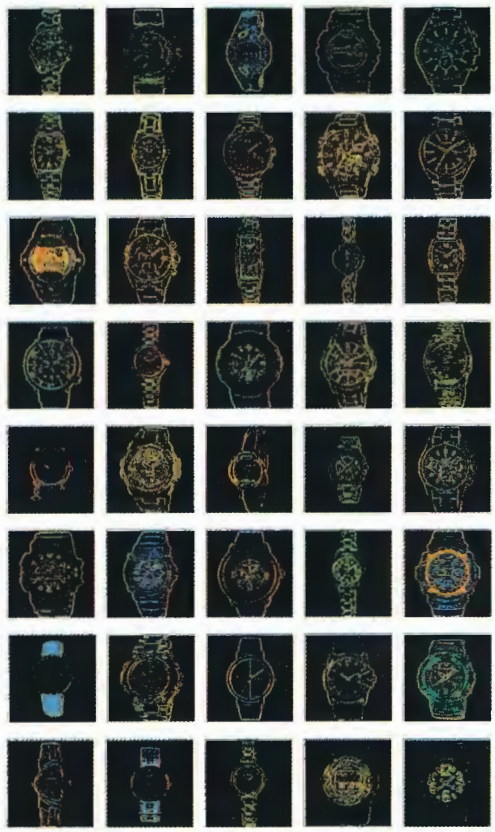
Grad-CAM

Grad-CAMは、画像の特徴抽出を行い、CNNが学習した概念を可視化する手法である。具体的には特徴マップのある位置に勾配の変化を与え、そのときに生じる出力の変化の大きさをもとに、あるクラスにとって重要な位置を特定する。そのため、出力層のクラスを指定することにより、入力画像におけるそのクラスの推定に重要な領域をヒートマップとして視覚的に表現が可能である。

可視化結果・考察

2.6節で述べた交差検証において最も推定誤差が小さかったCNNモデルに対し、(i)の処理を適用し、７枚のヒートマップ画像を取得した。その際、可視化する層は第２図中のConv2とした。得られた７枚のヒートマップに対して(ii)の処理を適用し、ヒートマップ統合画像を取得した。得られた画像の代表例を第５図に示す。

次に、得られた結果の傾向を把握するため、使用したデータセットにおける代表的なサンプルを



第５図　ヒートマップ統合画像例

選定した。その際、評価値の分布に対してAffinity Propagationによるクラスタリングを行った。クラスタリングの結果、「上品だ、可愛らしい、小さい」の印象に対して本データセットは16クラスタに分類でき、それぞれクラスタ内での代表的なサンプルを選定した。第６図に比較的高い印象値が付与されたクラスタの代表サンプルを、第７図に比較的低い印象値を付与されたクラスタにおける代表サンプルをそれぞれ示す。図中にはサンプル画像だけでなく、評価値の分布（gt）、推定値の分布（pred）、およびヒートマップ統合画像を併せて図示した。また、ヒートマップ統合画像の下にクラス番号に対応した色を記載した。

第６、７図から、腕時計の形状を保持した状態で印象に寄与する画像領域が可視化されていることが分かる。これは入力層に近い畳み込み層を用いたためと考えられる。

また、第６、７図のヒートマップ統合画像を比較すると、第６図では、高評価（印象値５－７）を示す箇所は腕時計の輪郭付近であり、低評価（同１－３）を示す箇所が腕時計のバンド付近であることがわかる。このことは、高評価の推定には輪郭が強く寄与しており、低評価の推定にはバンドが寄与していることを示唆している。すなわち、全体的な評価が高い腕時計画像では、高い評価を付ける人は輪郭や光沢を重視し、低い評価を付ける人はバンドを重視している可能性が考えられる。同様に第７図では、高評価（同５－７）を示す箇所は腕時計の輪郭付近であり、低評価（同１－３）を示す箇所は文字盤の文字付近であることがわかる。このことから、全体的な評価が低い腕時計画像では、高い評価を付ける人は輪郭を重視しており、低い評価を付ける人は文字盤の文字を重視している可能性が考えられる。

以上より、提案手法を用いることで人が画像から喚起される印象について、重視している箇所や評価のばらつきを空間的に説明できる可能性を示した。

第６図　高い印象値が付与された クラスタの代表サンプル

第７図　低い印象値を付与された クラスタにおける代表サンプル

おわりに

本稿では、プロダクト画像の印象推定において生じる「個人のばらつき」を包含した高精度な印象推定モデルの構築手法について、また、構築したモデルを用いて印象に寄与する画像領域の可視化する手法について紹介した。

画像と印象との関係性をモデル化するため、画像から抽出したスタイル特徴と、離散確率分布で表現された印象値を学習し、印象推定を行うCNN（印象推定モデル）を構築したことを確認した。その後、CNNにおける印象推定への寄与の度合いを可視化する手法（Grad-CAM）を利用して、構築したCNNにおいて印象に寄与する画像領域を可視化した。その結果、評価傾向の異なるサンプル間では、印象に寄与する画像領域が異なることが確認された。これにより印象評価における個人のばらつきを空間的に説明できる可能性を示した。

参考文献

1) 長田典子,「感性の指標化とプロダクトデザインへの応用」, 電子情報通信学会誌, 102(9), 873-880 (2019)
2) M. Nagamachi, "Kansei Engineering: a New Ergonomic Consumer-oriented Technology for Product Development", International Journal of industrial ergonomics, 15(1), 3-11 (1995)
3) 饗庭絵里子, 高松直也, 沼田晃佑, 柳田修太, 鈴木征一郎, 佐藤暢, 長田典子, 高田勝啓,「年代による感性空間の違い」, 日本感性工学会論文誌, 15(7), 677-685 (2016)
4) 井口征士, 猪田克美, 小林重順, 田辺新一, 長田典子, 中村敏,「感性情報処理」, 電子情報通信学会編ヒューマンコミュニケーション工学シリーズ (1994)
5) 飛谷謙介, 松本達也, 谿雄祐, 藤井宏樹, 長田典子,「素肌の質感表現における 印象と物理特性の関係性のモデル化」, 映像情報メディア学会誌, 71(11), 259-268 (2017)
6) N. Murray, L. Marchesotti, and F. Perronnin, "Ava: A Large-scale Database for Aesthetic Visual Analysis", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.2408-2415 (2012)
7) S. Kong, X. Shen, Z. L. Lin, R. Mech, and C. C. Fowlkes, "Photo Aesthetics Ranking Network with Attributes and Content Adaptation", in Proceedings of the European Conference on Computer Vision, 662-679 (2016)
8) J. Ren, X. Shen, Z. Lin, R. Mech, and D. J. Foran, "Personalized Image Aesthetics", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 638-647 (2017)
9) K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", in Proceedings of the IEEE Conference on Computer vision and pattern recognition, 770-778 (2016)
10) M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks", in Proceedings of the European Conference on Computer Vision, 818-833 (2014)
11) 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", in Proceedings of the IEEE International Conference on Computer Vision, 618-626 (2017)
12) N. Sunda, K. Tobitani, A. Takemoto, I. Tani, Y. Tani, T. Fujiwara, N. Nagata, and N. Morita, "Impression Estimation Model and Pattern Search System Based on Style Features and Kansei Metric", in Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, 1-2 (2018)
13) 鈴木秀通, 飛谷謙介, 橋本翔, 山田篤拓, 長田典子,「レビューテキストと画像を用いた機械学習によるプロダクトの感性指標構築」, 精密工学会誌, 85(12), 1143-1150 (2019)
14) L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2414-2423 (2016)
15) D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", arXiv preprint arXiv:1412.6980 (2014)
16) K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification", in Proceedings of the IEEE international conference on computer vision, pp.1026-1034 (2015)
17) X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks", in Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 315-323 (2011)
18) S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", arXiv preprint arXiv:1502.03167 (2015)
19) X. Jin, L. Wu, X. Li, S. Chen, S. Peng, J. Chi, S. Ge, C. Song, and G. Zhao, "Predicting Aesthetic Score Distribution through Cumulative Jensen-Shannon divergence", in Proceedings of the 32nd AAAI Conference on Artificial Intelligence (2018)
20) J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248-255 (2009)

【筆者紹介】

飛谷 謙介
関西学院大学　理工学部
感性価値創造インスティテュート

谷　伊織
関西学院大学　理工学部
感性価値創造インスティテュート

橋本　翔
関西学院大学　理工学部
感性価値創造インスティテュート

長田 典子
関西学院大学　理工学部
感性価値創造インスティテュート

〒669-1337　兵庫県三田市学園２－１
TEL：079-565-8300
E-mail：tobitani@kwansei.ac.jp