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
MVE2021-14(2021-09)

スタイル特徴を利用したCNNによる印象分布推定

大萩　優哉† 　飛谷　謙介†† 　谷　　伊織††† 橋本　　翔†††† 長田　典子†

† 関西学院大学大学院理工学研究科　〒669–1337 兵庫県三田市学園 2–1
†† 長崎県立大学情報システム学部　〒851–2195 長崎県西彼杵郡長与町まなび野 1–1–1
††† 神戸大学情報基盤センター　〒657–8501 兵庫県神戸市灘区六甲台町 1 – 1
†††† 西南学院大学商学部　〒814–8511 福岡県福岡市早良区西新 6–2–92
E-mail: †{Yuya-Ohagi,nagata}@kwansei.ac.jp, ††tobitani@sun.ac.jp, †††iori_tani@penguin.kobe-u.ac.jp,
††††s-hashimoto@seinan-gu.ac.jp

あらまし　本研究では，深層学習モデルを使用し，プロダクトの画像の印象評価時に生じる個人差を考慮した印象分布推定モデルの構築手法を提案する．まず，プロダクトの画像から抽出される視覚的印象と強い関係があると示唆される特徴量（スタイル特徴）と画像の印象分布の両方を使用してプロダクトの印象分布を推定する CNN を構築した．その後，CNN の印象分布推定に寄与する画像領域を Grad-CAM を用いて可視化し，人が評価時に重視する画像領域を心理実験で取得し，互いの画像を比較することで画像領域の類似性を確認した．
キーワード　感性工学，視覚的印象，印象推定，可視化，深層学習

A CNN model Using Neural Style Features for Predicting Aesthetic Impressions Score Distribution

Yuya OHAGI†, Kensuke TOBITANI††, Iori TANI†††, Sho HASHIMOTO††††, and Noriko NAGATA†

† School of Science and Technology, Kwansei Gakuin University　2–1 Gakuen, Sanda-shi, Hyogo, 669–1337 Japan
†† Faculty of Information Systems, University of Nagasaki　1–1–1 Manabino, Nagayo-cho, Nishi-Sonogi-gun, Nagasaki, 851–2195 Japan
††† Information Science and Technology Center, Kobe University　1–1 Rokkodai-cho, Nada-ku, Kobe, 657–8501 Japan
†††† Faculty of Commerce, Seinan Gakuin University　6–2–92 Nishijin, Sawara-ku, Fukuoka, 814–8511 Japan
E-mail: †{Yuya-Ohagi,nagata}@kwansei.ac.jp, ††tobitani@sun.ac.jp, †††iori_tani@penguin.kobe-u.ac.jp,
††††s-hashimoto@seinan-gu.ac.jp

Abstract　In this study, we propose a method for predicting the probability distribution of aesthetic impression scores considering individual differences in impression evaluations using a deep neural network. We adopted neural style features, which potentially have relationships with visual impressions as explanatory variables. Then, we constructed a convolutional neural network (CNN) that estimated the probability distribution of impression scores based on product images. Next, we visualized attention maps that represented image areas that contribute to impression scores by using Grad-CAM. We also conducted an impression evaluation experiment to relate individual impression scores to the image areas that each participant considered important. Finally, we confirmed the similarity among the image areas by comparing the attention maps and the experimental results.
Key words　kansei (affective) engineering, visual impression, impression estimation, visualization, deep learning

1. 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．また，ユーザは高級さといった感性的なニーズも重要なものと考えている [1]．このような感性的ニーズを把握する場合，信頼性が高く有効な方法論として感性工学のアプローチ等があげられるが，高精度な感性指標を構築するためには心理実験によるデータの取得とその分析にかかる人的

及び時間的な負荷が高いという問題がある．そこで，深層学習による技術の自動化が注目されている．その一例として，パッケージデザインの分野ではデザイナがより効率的にデザインの良し悪しをスコアやヒートマップから判断できる手法が提案されている [2]．しかしながら，プロダクトごとに個人が重視するデザインを反映する仕組みについては十分に考慮されていない．

そこで本研究では，視覚的印象に着目し，プロダクトの画像の印象評価時に生じる個人差を考慮した印象分布推定モデルの構築を行うことを目的とする．

2. 先行研究

プロダクトの画像と視覚的印象とのモデル化では，画像特徴量として色，光沢感や表面粗さ，形状などが使用されてきた [3]．近年では，深層学習の発展に伴い，一般物体認識で高い成果を出している CNN から抽出される画像特徴量も用いられている．特に VGG-19 [4] を利用して Gatys らが提案した画風変換アルゴリズムの中で使われたスタイル特徴は画像中のマルチスケールな色やパターン等の詳細な見た目を表現しており，後にその特徴量を利用しテクスチャと印象をモデル化した研究から視覚的印象と強い関係があることが示唆されている [5] [6]．しかしながら，プロダクトの画像と視覚的印象を，スタイル特徴を用いてモデル化した研究は十分になされていない．

一方で，深層学習の分野では，判断根拠を明らかにする研究が盛んに行われている．特に，CNN においては，モデルが学習した概念を可視化する Grad-CAM という手法が提案されている [7]．Grad-CAM は CNN において可視化対象の畳み込み層の各領域が出力に寄与する度合いを疑似的にヒートマップとして表現する手法である．

本研究では，プロダクトの画像の印象とスタイル特徴が強い関係があると考え，その特徴量を説明変数とした印象分布推定モデルを構築する．更に，Grad-CAM を利用してモデルが重視したプロダクトの画像領域を明らかにし，人が実際に重視する箇所との類似性からモデルの妥当性を確認する．

3. スタイル特徴を用いた CNN

本研究では入力データである画像から学習済み VGG-19 を通して抽出したスタイル特徴を利用して印象分布推定モデルを構築する．先行研究において，Sunda et al. は印象推定モデルの構築の際に各 Pooling 層から抽出されたスタイル特徴を使用しており，この特徴量が感性情報の中でも視覚的印象との関わりが強い可能性があることを示した [6]．そのため，この特徴量を利用することで高精度な印象推定が実現できると考えられる．その際，深い Pooling 層から抽出されたスタイル特徴ほど高次元であり，データセットのサンプル数が少ない場合は学習の際に過学習することが予想される．そこで本研究では，最初の Pooling 層から抽出されるスタイル特徴を用いる．

3.1 スタイル特徴の概要

スタイル特徴は，一般物体認識に用いられる CNN である VGG-19 の中間層 l から出力される特徴マップの相互相関行列 G^l である [5]．G^l の次元数は l 層の特徴マップの数を N_l とすると $N_l \times N_l$ であり，その成分 G^l_{ij} は，l 層の特徴マップのインデックスである i と j，及びベクトル化された n 番目の特徴マップの m 成分 F^l_{nm} を用いて式 1 のように表される．特徴次元数は，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．本研究では，64 × 64 の 4096 次元のスタイル特徴のみを利用する．

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk} \tag{1}$$

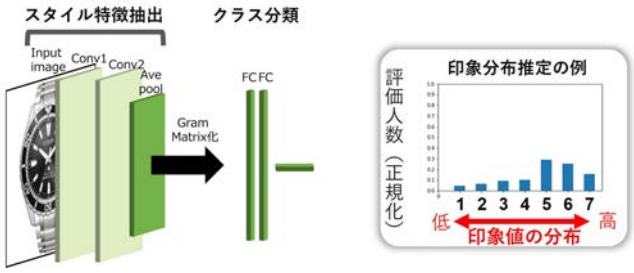

図 1　CNN の構造
Fig. 1　CNN architecture

表 1　CNN の構造の詳細．入出力サイズ：$height \times width \times channels$
Table 1　Detail of CNN architecture. The input and output sizes are denoted as $height \times width \times channels$

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		

3.2 印象分布推定を行う CNN の構築

本研究で扱う CNN は，画像を入力データとし，スタイル特徴を抽出し，二層の全結合層を付加し，出力層でクラス分類を行う．その際，クラス分類は印象の高低と対応し，クラス番号と印象値の段階が一致している．構築した CNN の構造及び詳細をそれぞれ図 1，表 1 に示す．図 1 における Conv1 及び Conv2 の層では VGG-19（ImageNet [8] にて学習済み）の重みを用い，学習による重みの更新は行わない．また，表 1 において，Ave pool の層の出力から相互相関行列の抽出を行う．そのうえで，10-fold 交差検証を行い，推定誤差が最小のモデルを 4 章で述べる画像領域の可視化に用いる．

CNN の学習における最適化手法として adam [9] を使用し，バッチサイズは 128 とする．全結合層の重みの初期値には He の初期値 [10] を適用し，学習率の初期値は 1.0×10^{-6} とする．活性化関数は Rectified Linear Unit（ReLU）関数 [11] を使用し，全結合層では Batch Normalization [12] を適用する．出力層では softmax 関数を使用し，損失関数にはクロスエントロピー誤差

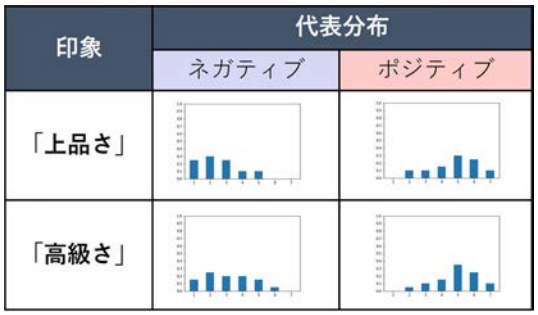

図 2　代 表 分 布
Fig. 2　Representative distribution

を使用する．加えて，推定誤差の算出には MSE (Mean Squared Error) を使用する．

3.3　対象データセット

本研究では，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] を用いる．その理由としては，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しやすく，結果の有用性・妥当性を検証しやすいためである．

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」（以下，「上品さ」）及び「高級さ，重い，重厚だ」（以下，「高級さ」）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 1-7 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である「クラウドワークス」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．

3.4　印象分布推定結果

本研究では，印象値の分布の傾向を判断基準として CNN の推定精度を算出するために次の (i)〜(iv) の処理を行う．(i). データセット全体の印象値の分布（以下，gt）を対象に ward 法を用いて，低評価寄りの分布クラスタ（以下，ネガティブ）と高評価寄りの分布クラスタ（以下，ポジティブ）の 2 クラスタに分割する．(ii). 各クラスタにおいて，ユークリッド距離を使用し，距離が最小のサンプルを代表分布とする．(iii). 推定された分布と 2 つのクラスタの代表分布のユークリッド距離を算出し，距離が最小のクラスタを選択する．(iv). 各クラスタにおいて gt のクラスタと (iii) で選択されたクラスタの一致率を算出し，これを推定精度とする．以上の処理で決定した代表分布，CNN の推定精度を算出した結果をそれぞれ，図 2，表 2 に示す．

表 2　推 定 精 度
Table 2　Estimation accuracy

印象	精度 [%] (サンプル数/総サンプル数)		
	ネガティブ	ポジティブ	平均
上品さ	68.6 (48/70)	82.3 (107/130)	77.5 (155/200)
高級さ	66.0 (64/97)	81.6 (84/103)	74.0 (148/200)

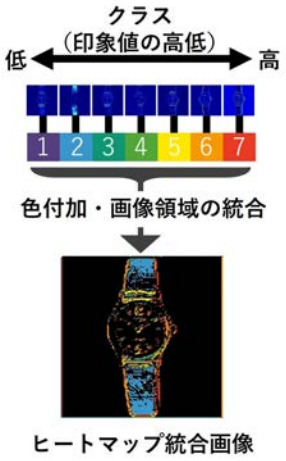

図 3　可視化フロー
Fig. 3　Visualization flow

結果からポジティブ，ネガティブを合わせて多数のサンプルで評価傾向を捉えた推定が行えることが確認できた．

4.　印象分布推定に寄与する画像領域の可視化

CNN モデルに対し，印象分布推定に寄与する画像領域を可視化し，人が実際に重視する箇所と比較することでモデルの妥当性を確認する．

4.1　可視化フロー

本研究では，印象分布推定に寄与する画像領域の可視化を行うために 2 つの処理を行う．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．可視化フローを図 3 に示す．

（i）ではクラス分類における各クラス（印象の高低）に寄与する画像領域を Grad-CAM により可視化を行い，ラベル数分のヒートマップを取得する．(ii) では（i）で得られた全てのヒートマップを用いて画素毎に比較を行い，最も寄与度の高いクラスに対応する色（図 3 のクラス番号の色）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，背景画像の推定値への影響を除いて可視化結果の傾向を確認するために，ヒートマップ統合画像は背景の寄与度の値以下の寄与度を持つ画素は黒色とした．

4.2　可視化処理

CNN モデルに対し，(i) の処理を適用し，各画像に対して，教師データの 7 クラスに基づく 7 枚のヒートマップが得られた．その際，可視化する層は図 1 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

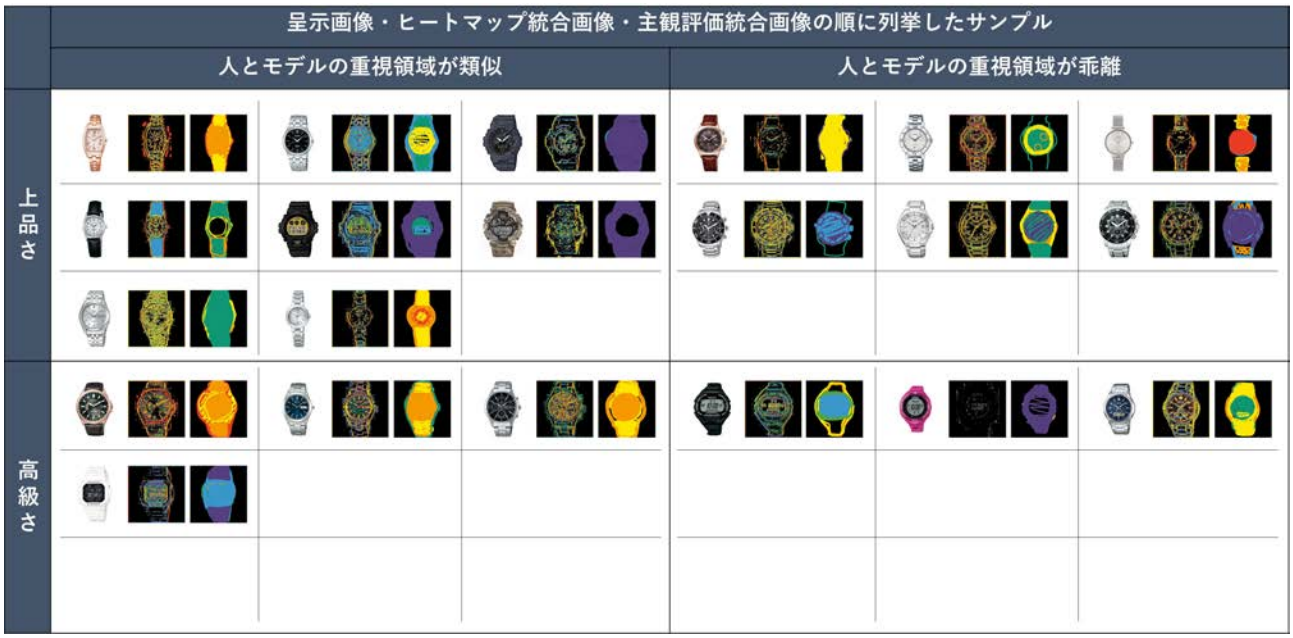

図 4　実験結果（各サンプルは左から，呈示画像，ヒートマップ統合画像，主観評価統合画像）
Fig. 4　Experimental results. The image of each sample is from left to right: presented image, integrated image of the attention maps, and integrated image of the subjective evaluations.

プ統合画像を取得した．

4.3　検 証 実 験

印象分布推定に寄与する画像領域と人が重視する画像領域を比較するため，印象評価実験を行った．

4.3.1　実験刺激の選定

ヒートマップ統合画像に対してクラスタ分析を行い，網羅的で代表的なサンプルを実験刺激として選定した．選定の際，正しく推定されなかったサンプルは除いた．クラスタ分析の際は，ヒートマップ統合画像からコンテンツ特徴量を抽出し，その特徴量に対して ward 法を用いた．コンテンツ特徴量とは，一般物体認識に必要な特徴量であり，スタイル特徴と対比される特徴量である．クラスタ数は，集中度孤立度分析から「上品さ」が 8，「高級さ」が 7 とした．更に，各クラスタのサンプルの gt をポジティブ・ネガティブで分割し，各クラスタ重心のサンプルを選定し，最終的に実験に用いる刺激数は「上品さ」が 15，「高級さ」が 12 とした．

4.3.2　実験参加者

実験参加者は，大学生及び大学院生 10 名（男子学生 5 名，女子学生 5 名）であった．

4.3.3　実験手続き

実験試行が始まると，画面上に実験刺激と評価入力欄が呈示される．実験参加者には，呈示された実験刺激を観察してもらい，着目する印象がどの程度あてはまるかについて「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」の評価尺度からなる 7 件法で回答を求めた．また，印象評価時に重視した箇所を回答及び図示するように求めた．

4.3.4　実験結果・考察

評価の際に人が重視した画像領域と，ヒートマップ統合画像を比較するため，実験で得られた画像領域から人の疑似的な重視領域をヒートマップで示したものとして，「主観評価統合画像」を作成した．主観評価統合画像とは，実験刺激ごとに，評価において重視した画像領域にあたる画素の画素値を対応する評価値で置換し，その結果得られた実験参加者分の画像において画素位置ごとに最頻値を取得し，評価値に対応する色を付加した画像である．

また，データセットに付与された評価値を付与した人と今回の実験参加者の属性が異なることが考えらえるため，ヒートマップ統合画像と主観評価統合画像を比較するサンプルは共通の評価傾向を持つことが望ましい．そこで推定精度を算出した際と同様に，gt のクラスタと実験で得られた評価値の分布のクラスタが一致するサンプルを選定した．選定されたサンプル数は「上品さ」が 14，「高級さ」が 7 であった．選定されたサンプルについて，呈示画像，ヒートマップ統合画像，主観評価統合画像を図 4 に示す．

ヒートマップ統合画像と主観評価統合画像の比較を行うことで互いの類似性を確認した．人とモデルの重視領域間の全体的な傾向として，シンプルな文字盤やフレームや小さい文字を含むといった特徴を持つサンプルは類似する傾向がみられ，複雑な文字盤やフレームや大きい文字を含むといった特徴を持つサンプルは乖離する傾向がみられた．以上のことから大域的なパターンの印象は捉えられていないが，局所的なパターンの印象を捉えた印象分布推定が行える可能性が示唆された．

5.　お わ り に

本研究では，プロダクトの画像の印象評価時に生じる個人差を考慮した印象分布推定モデルの構築手法を提案した．まず，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

スタイル特徴を用い，離散確率分布で表現された印象値を学習し，印象分布推定を行う CNN を構築した．次に，CNN の印象分布推定に寄与する画像領域を可視化した．その後，CNN と人が重視する画像領域の類似性を確認するため，検証実験を行い，人が重視する画像領域を取得した．その結果，局所的なパターンの印象においては人と CNN の重視領域が類似する傾向がみられ，人が重視する箇所を捉えた印象分布推定が行える可能性を示した．

今後の課題として，大域的なパターンの印象を捉えた印象分布推定が挙げられる．そのため，より高次元の特徴量等にも着目したモデル構築の検討を行う．

文　　献

[1] A. Takemoto，K. Tobitani，Y. Tani，T. Fujiwara，Y. Yamazaki，N. Nagata，Texture synthesis with desired Visual impressions using deep correlation feature，2019 IEEE International Conference on Consumer Electronics (ICCE)，pp.739-740，2019．

[2] 夏博恵，坂元英樹，汪雪テイ，山崎俊彦，深層学習を用いたパッケージデザインの好意度予測，人工知能学会全国大会論文集 第34 回全国大会 (2020)，1M3GS1302，2020．

[3] 丹羽志門，青山祥貴，数藤恭子，谷口行信，加藤俊一，商品写真から受ける印象と画像特徴の関係のモデル化，研究報告ヒューマンコンピュータインタラクション (HCI)，vol.2013，no.24，pp.1-4，2013．

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

[5] L. A. Gatys，A. S. Ecker，and M. Bethge，Image style transfer using convolutional neural networks，Proceedings of the IEEE conference on computer vision and pattern recognition，pp.2414-2423，2016．

[6] 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，Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology，P3-09，2018．

[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，Proceedings of the IEEE international conference on computer vision，pp.618-626，2017．

[8] J. Deng，W. Dong，R. Socher，L. J. Li，K. Li，and L. Fei-Fei，Imagenet: A large-scale hierarchical image database，2009 IEEE conference on computer vision and pattern recognition，pp.248-255，2009．

[9] D. P. Kingma，and J. Ba，Adam: A method for stochastic optimization，arXiv preprint arXiv:1412.6980，2014．

[10] K. He，X. Zhang，S. Ren，and J. Sun，Delving deep into rectifiers: Surpassing human-level performance on imagenet classification，Proceedings of the IEEE international conference on computer vision，pp.1026-1034，2015．

[11] X. Glorot，B. Antoine，and Y. Bengio，Deep sparse rectifier neural networks，Proceedings of the fourteenth international conference on artificial intelligence and statistics，pp.315-323，2011．

[12] S. Ioffe，and C. Szegedy，Batch normalization: Accelerating deep network training by reducing internal covariate shift，International conference on machine learning，pp.448-456，2015．

[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，vol.85，no.12，pp.1143-1150，2019．