

# 深層学習による質感情報に注目したテクスチャ画像の分類

## Texture image classification focusing on feel of materials

### using deep learning

東畑 和希\*  
Kazuki Higashihata

青木 茂樹\*  
Shigeki Aoki

宮本 貴朗\*  
Takao Miyamoto

## 1 はじめに

人間は物体を視認する際に、物体の種類や名称だけではなく物体の状態や質感などの情報も同時に認識している。同じ物体であっても質感によって異なる印象を受け取る場合もある。そして、質感情報をオノマトペで表現して認識している。オノマトペとは物事の状態などを象徴的に表した単語であり、質感を表す単語も含まれている。

近年、深層学習を利用した物体認識に関する研究が盛んに行われている。一般的な深層学習による物体認識手法では、物体の構造や形状から種類や名称などを認識できるものの、物体の質感情報の分類には主眼が置かれていなかった。

深層学習を利用した質感情報の認識に関する代表的な研究として、文献 [1] が挙げられる。この研究ではアート画像におけるスタイル情報の認識精度向上を目的として、CNN(Convolutional Neural Network) の中間層から抽出した特徴行列をスタイル情報として抽出し、抽出したスタイル情報を SVM(Support Vector Machine) で学習し画像を識別する手法を提案している。この手法では、特徴行列を利用して分類するよりも高精度に識別できているものの、精度が 75% 程度であり精度向上の課題があると考えられる。本手法では、質感情報を表すオノマトペに基づいてテクスチャ画像をニューラルネットワークにより高精度に分類する手法を提案する。

本手法では、画像をオノマトペに基づいて分類できるため、質感に関する印象の影響が大きいと考えられるファッションやインテリア分野での応用を期待できる。

## 2 提案手法

### 2.1 画像データセットの作成

まず、画像データセットを作成する。本研究では公開テクスチャ画像データセットである DTD(Describable Textures Dataset)[2] を利用する。DTD は 47 種類のテクスチャラベルで構成されており、各ラベル毎に 120 枚の画像が存在する大規模な画像データセットである。DTD のラベルには「waffled」や、「crystalline」、



図 1. 画像データセットの例

「knitted」のように物体の素材を表すラベルが多く、質感情報を表すラベルは少ない。そこで本研究では質感情報を表す、オノマトペを利用してデータセットを再構築する。各画像がどのオノマトペに属するかをアンケートで調査することにより新たにラベルを付けて、図 1 に示すような新たなテクスチャ画像データセットを構成する。

### 2.2 質感特徴ベクトルの抽出

一般的な CNN を利用した手法では画像の構造的な特徴を重視し画像分類を行っている。本研究では、質感情報に注目するために文献 [1] のスタイル情報を応用する。文献 [1] の手法では、CNN(VGG16[3]) の中間層から抽出した特徴行列とその転置行列の積が入力画像の質感情報を表しているという調査結果 [4] に基づいて、算出された行列の積を特徴ベクトルに変換する手法を提案している。

まず、VGG16 に画像を 1 枚入力し、特定の間層  $l$  から特徴行列  $F^l$  を抽出する。次に、抽出した特徴行列  $F^l$  に次式を適用して質感に関する特徴行列  $G^l$  を算出する。

$$G^l = F^l(F^l)^T \quad (1)$$

特徴行列  $G^l$  は対称行列となっているため、次式のようになら三角成分のみを抽出する。

$$V^l = (G_{1,1}^l, G_{1,2}^l, \dots, G_{1,n}^l, G_{2,2}^l, \dots, G_{n,n}^l) \quad (2)$$

その後、 $V^l$  を最小値が 0、最大値が 1 となるように正規化し、質感特徴ベクトル  $V_{norm}^l$  とする。

### 2.3 画像分類

全結合層のみで構成されたニューラルネットワークによる教師あり学習によって質感特徴ベクトルを分類する。中間層は 3 層から構成し、各層では ReLU 関数を適

\* 大阪府立大学大学院人間社会システム科学研究科 Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University

表 1. 本手法と VGG16\_fc との精度比較

	本手法	VGG16_fc
accuracy	0.9545	0.8862

用し、出力層ではソフトマックス関数を利用する。前節で抽出した質感特徴ベクトル  $V_{norm}^l$  と対応するラベルをニューラルネットワークに入力して学習する。分類は、学習時と同様に画像から質感特徴ベクトル  $V_{norm}^l$  を抽出して、学習したニューラルネットワークに入力した時の、出力に基づいて行う。ここで、複数の質感を有する物体も存在すると考えられるため、最大の確率を予測ラベルとするのではなく、確率が閾値以上のラベル全てを予測ラベルとする。したがって 1 つの入力データに対して複数のラベルが予測される場合もある。

### 3 実験

#### 3.1 実験環境

実験では、オノマトペとして「がさがさ」、「きらきら」、「くしゅくしゅ」、「すべすべ」、「もふもふ」、「つるつる」の 6 種類を利用した。画像のラベル付けは、ブラウザに表示された DTD の各画像に対して、上記のどのオノマトペに属するか、もしくは「該当しない」かの 7 つの選択肢から 1 つをユーザーに選択させて、各画像で最も多いラベルを新たなラベルとして付与した。ラベル付けには 12 人が参加し 1 人当たり 360 枚を担当した。集計の結果、計 1326 枚にオノマトペのラベル付き画像データセットを作成できた。

データセットの各画像に対して左右反転、90 度、180 度、270 度回転を行い、計 10608 枚に水増しを行った後、訓練データ 5941 枚、検証データ 1485 枚、テストデータ 3182 枚に分割した。

質感特徴ベクトルは ImageNet で学習済みの VGG16 の block3\_conv1(56 × 56 × 256) 層から抽出して作成した。画像分類に使用するニューラルネットワークのパラメータはエポック数を 100、バッチサイズを 256、予測ラベルの閾値は 0.8 に設定した。また比較実験として、VGG16 のファインチューニングのみでのクラス分類も行った。

#### 3.2 実験結果

実験結果を表 1 に示す。実験の結果、accuracy:0.9545 となり本手法の有効性を確認できた。また、VGG16 でファインチューニングして分類した結果(表 1 の VGG16\_fc) と比べ、本手法における分類の方が高精度であることを確認することができた。

#### 3.3 考察

正しく認識できた画像例を図 2 に示す。「きらきら」は、光が反射している部分とそれ以外の部分での明度の差がはっきりしている画像や、全体的に彩度が高い画像が多かった。「つるつる」も、「きらきら」と同様に光の反射による明度の差がはっきりしている画像が多かった。しかし「つるつる」は、彩度の高い画像が「きらきら」と比較して少なかった。このような明度や彩度の表現の差異を学習できたために、正しく認識できたと考えられる。



図 2. 正しく認識された画像例

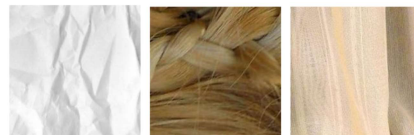


図 3. すべすべと誤認識された画像

誤認識した画像例を図 3 に示す。図 3 は、全て「くしゅくしゅ」でラベル付けされた画像を「すべすべ」であると誤認識してしまった画像例である。誤認識した画像の共通点として、全体的に白いことが挙げられる。このことから、画像全体に明度の差が余り無く白をベースとした色が広がっている特徴を「すべすべ」として学習したのではないかと考えられる。「すべすべ」の画像データセットには他のラベル画像と比較しその様な特徴の画像が多く存在していた。これらの結果から、質感特徴ベクトルをニューラルネットワークで学習した際に、全体的な色に対して注目する傾向があったのではないかと推察される。今回の実験では、VGG16 の block3\_conv1 から質感特徴ベクトルを抽出したが、どの中間層から抽出するかによって抽出できる特徴が変わると考えられる。今後、本研究に最適な中間層を検討したいと考えている。

#### まとめ

本稿では、オノマトペに注目したテクスチャ画像データセットから質感特徴ベクトルを抽出し、ニューラルネットワークを用いてテクスチャ画像を分類する手法を提案した。VGG16 をファインチューニングした手法と比較し本手法の有効性を確認することができた。一方、本手法による分類では画像の色に関する情報に大きく影響される傾向があるため、今後の課題としては色情報に影響されない特徴を検討することなどが挙げられる。

### 4 参考文献

- [1] Matsuo, Yanai: CNN-based Style Vector for Style Image Retrieval, Proceedings of the 2016 ACM International Conference on Multimedia Retrieval, pp.309-312, 2016.
- [2] Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A.: Describing textures in the wild, In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2014.
- [3] keran Simonyan, Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition, In Proc. of International Conference on Learning Representation, 2015.
- [4] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In Proc. of IEEE Computer Vision and Pattern Recognition, 2016.