深層学習を用いたモチーフ分類にもとづく小袖屏風画像の特徴分析

田村誠悟・濱上知樹 （横浜国立大学大学院工学府）

デジタルアーカイブの高度利活用をめざした取り組みとして，高精細画像からの知識の抽出と歴史研究支援システムの構築を進めている．その一環として，小袖屏風画像に含まれる様々なモチーフを深層学習によって自動抽出・分類し，その構造的共通点によって意味的類似性を定量的に発見する手法を提案する．102 隻の小袖屏風画像を分析した結果，抽象的な意匠の類似性によってクラスが構成できることが明らかとなった．

A feature classification from Kosode-byobu images based on motif analysis by the deep learning

Tamura, Seigo / Hamagami, Tomoki (Yokohama National University)

As an effort for advanced utilizations of digital archives, the method of knowledge retrieval from high definition images and an aid system for studies of literature and history are developing. As a part of it, this study proposes a new method to quantitatively evaluate semantic similarity among Kosode-byobu images, based on the motif analysis by the deep learning. The results of analysis of the 103 images shows abstract image patterns can be classified well by using this method.

１．まえがき

近年，歴史資料のデジタルアーカイブ化が急速に進んでいる．特に画像データは ICT の発展とともにその質・量ともに飛躍的な伸びをみせている．同時にその高度な利活用方法について，さまざまな試みが開始されている．

蓄積された莫大な画像１次データを研究に活用しようとした場合，解釈可能なメタデータの付与が問題になる．ここでメタデータとは，データそのものを説明するデータである．１次データの精度や量が増えることにより，メタデータを人の解釈によって付与することは困難になっていく．

この課題に対し．大量の画像データの中から，機械学習によって画像中に含まれる様々な意味を抽出・分類することが検討されている．機械学習を用いた画像からの特徴抽出は，一般物体認識の分野で多くの研究の蓄積があり，これらの既存手法を応用した例もある．しかし，それぞれの資料固有の特徴に注目した手法が多く，汎用的な手法の確立には至っていない．

歴史資料の中でも創作物の多くは，様式や意味に沿った抽象的な構造を有している．その中でも．モチーフ（文様）は，様式に代表される単位であり，様々な資料で繰り返し使われるデザインの基底になりうる．これらの基底が自動抽出できれば，大量の画像データに対するメタデータの付与や特徴の発見新たな知的構造の発見を通して，歴史研究支援や高度な展示などに貢献することをめざしている(1).

本稿では，この国立歴史民俗博物館蔵の小袖屏風（野村コレクション(2)）を対象として，屏風画像に含まれるモチーフの定量的抽出と特徴分析を，深層学習(Deep Learning)を用いて行う手法について検討した．

２．小袖屏風デジタルアーカイブ

小袖屏風とは小袖を二曲一隻に仕立てた屏風である．国立歴史民俗博物館には，野村コレクションの一部である約 100 余隻が収蔵されている．辻が花，縫箔，慶長小袖，寛文小袖、元禄小袖，友禅染などの小袖の諸相を網羅しており，芸術品としてはもちろん，染織史研究にとっても貴重な資料価値がある． 本研究では，この小袖屏風の高精細デジタルデータ（原寸に対し約 200dpi）からなる小袖屏風デジタルアーカイブを対象としている．図１に，国立歴史民俗博物館に所蔵されている野村コレクションの例を示す．

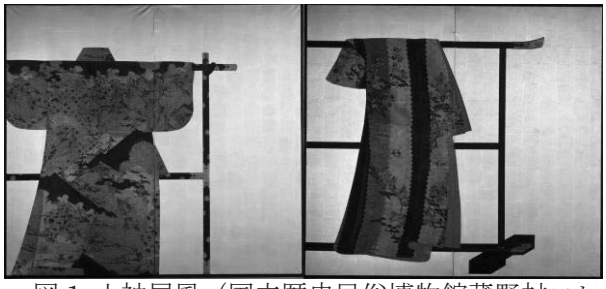

図１ 小袖屏風（国立歴史民俗博物館蔵野村コレクション）

ⓒ 2015 Information Processing Society of Japan

小袖の絵柄は様々なメタ情報を含んでいる．分類上の特質（文様，地色，地質，諸技法）のほかに，メタファー，背景にあるストーリー，マクロからミクロに至る構図上の特徴などの非言語的な情報も含んでいる．これらの小袖屏風のメタ情報が与える印象の類似性や，陽に現れない部分の共通性などが抽出できれば，小袖屏風に含まれる隠れたセマンティクスや新たな構造の発見につながることが期待できる．

また，小袖屏風の資料分析においては，小袖表面の柄や織，文様等の様々な特徴をもとに，他資料間との類似性や連続性に着目した分析が行われる．屏風上に貼られた小袖は，元来の様子のままとは限らず，断片からの再構成や創作となっている場合も多い．これらの制作過程までを含めた資料をまたぐ情報の構造が抽出できれば，これまでにない展示，検索等への応用が期待できる．

本研究においては，特に小袖屏風にみられるさまざまな文様に注目する．着物にみられる文様は，装飾としての目的だけではなく，季節や吉祥を意味する意匠として，さまざまなモチーフが用いられている．これらのモチーフの構造からなる類型化により，小袖屏風の解釈を支援する方法を提案する．

３．機械学習によるモチーフの抽出

機械学習を用いたモチーフ抽出の方法としては，従来 SIFT (Scale Invariant Feature Transform) や SURF(Speeded Up Robust Features)のように画像中の濃淡勾配を利用した特徴量が用いられている.また，これらの特徴をベクトルで表現し，その集まり(Bag)によって，画像の構造的特徴をとらえる方法(Bag of Features)がしばしば用いられる.しかし，これらの手法を小袖屏風のモチーフのように，濃淡のみで表現されない抽象的意匠に用いることは適切ではない．

この問題に対し，筆者らはこれまで一般物体認識空間における文様の分類方法[3]を提案してきた．この手法は，画像中の特徴的な点を階層的なクラスタリングを用いてセグメントを作成し，各セグメントを一般物体認識器として深層学習によって学習された識別器を用いて評価することで，101 次元の一般物体認識空間で表現する．この方法により，実体とは異なる抽象的なモチーフであっても，「～のように見える」という例えによって，抽象的な特徴分類をすることが可能になった．図 2 に，この手法の概要を示す．

既提案方法では，文様の類型化について一定の効果が確認できたものの，学習対象となっている一般物体が，モチーフを表現する上で必ずしも適した基底ではないという問題が残されていた．

そこで，一般物体認識を基底とするのではなく，深層学習の中間学習層を用いて，モチーフの基底を構成する試みを行う．さらにそのクラスタリングと分布によって，小袖屏風全体の特徴分析を定量的に行う手法へ発展させる．

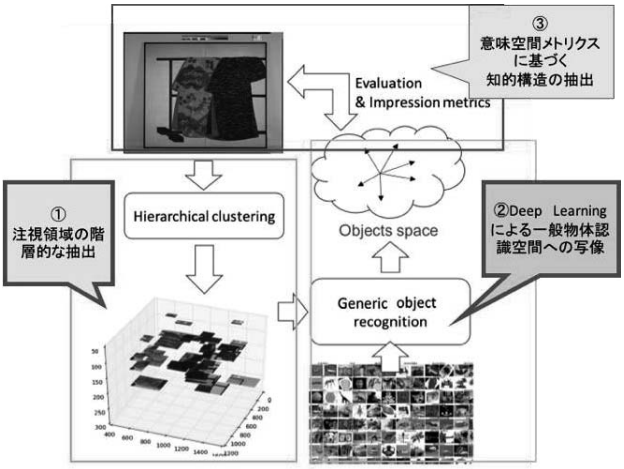

図 2　一般物体認識器を用いたモチーフ分類[3]

４．Caffe/DeCAF によるモチーフ分析

4.1　Caffe

Caffe とは，BVLC で開発された深層学習のライブラリおよびツールである[4]．多くのアプリケーションでの実績があり，特に 1000 カテゴリからなる特に ImageNet データベースを用いた学習済みのネットワークは，高い一般物体認識能力を有していることが知られている．

図 3 に Caffe による深層学習モデルである，CNN(Convolutional Neural Network)の構造を示す．ここでは 5 段の畳み込み(conv)層と，プーリング(pool)層を経て，Full-connected の結果である 4096 次元の特徴ベクトルがつくられる．この 4096 次元のベクトルに，入力画像に対する基底成分が現れる．

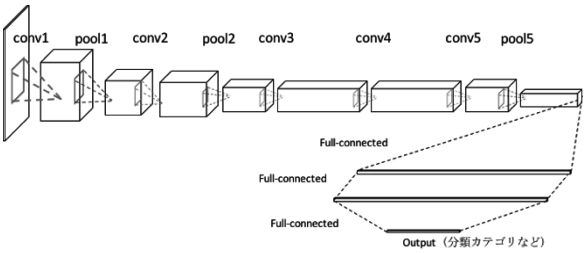

図 3　Caffe による CNN の実装

畳み込み層（図 4）では，画像の局所領域において式(1)に示す畳み込みを行い，情報の圧縮を図る．この畳み込みを複数のフィルタを用いて行う．

$$h_j^n = \phi\left(\sum_{k=1}^{K} h_k^{n-1} * w_k^n\right) + b_k \quad \cdots\cdots \quad (1)$$

ⓒ 2015 Information Processing Society of Japan

ここで, ϕは活性化関数, wはフィルタ, Kは$N-1$層目の特徴マップ，bはバイアスである.

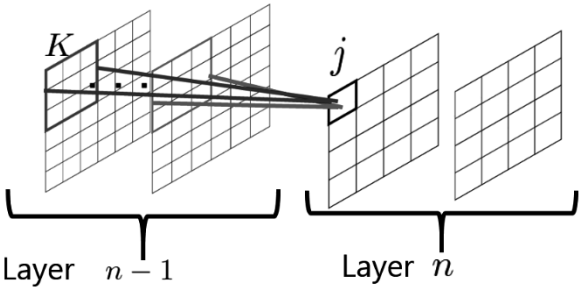

図 4　　畳み込み層

つぎに，プーリング層（図 5）では式(2)に示すように，局所領域の情報を一部捨て,情報を間引くことで画像中の物体位置の違いを吸収している. これらのネットワークを多段に重ねることにより, 中間ノードに画像を構成する基底のイメージが抽出される.

$$h_j^n(x,y) = \max h_j^{n-1}(\bar{x},\bar{y}) \cdots\cdots \quad (2)$$

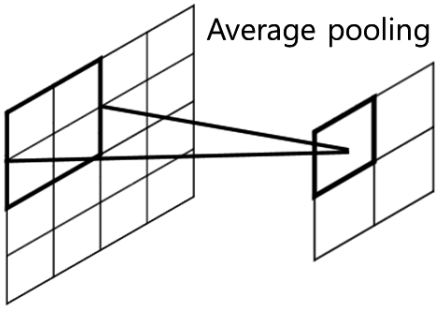

図 5　プーリング層

4.2　DeCAF

学習時には，output 層に与えるカテゴリベクトルと，入力層との間を，誤差逆伝播法(Back propagation)を用いて学習する.

具体的には，ImageNet のデータベースを学習し，あらかじめ中間層に基底ノードを作成する. このように, Caffe のネットワークの中間層から抽出される 4096 次元の特徴空間は DeCAF (Deep Convolutional Activation Feature)[5]と呼ばれている. DeCAF には，一般物体認識に有用な基本的特徴を有しており, 様々な画像処理において有効性が確認されている.

4.3 小袖屏風画像の BoF(Bag of Features)

本研究では, 以下の手順に従って小袖屏風画像の小領域から DeCAF を抽出し，その空間上の処理によって特徴を分類する.

1. 階層的画像セグメンテーション
   すべての領域を対象としたモチーフの分析を行うために，102 隻の屏風画像データをそれぞれ 1200 分割した部分画像を最小単位として用いる. 最小部分領域の実寸は，およそ 5×5cm である. さらに，最小部分領域を 4 つ合わせた 320 領域で部分画像をつくる. 同様にして，80 分割，20 分割, 6 分割の部分領域をつくり, 計 1620 領域/屏風のデータ，合計 165852 領域の画像を得る. 以降これをセグメントと呼ぶ. なお，これらのセグメント画像はすべて227×227にリサイズされており, 端の画像は拡大することですべてのセグメントのサイズを正規化する.
2. DeCAF によるベクトル化
   セグメント画像をすべてを ImageNet で学習済みの Caffe モデルを用いて，DeCAF の 4096 次/セグメントの特徴ベクトルに変換する.
3. クラスタリング
   DeCAF のベクトルを k-means 法を用いて 100 のクラスに圧縮する. ここで クラス数は各クラスにモチーフが共通して分類されるよう経験的に設定した数値である.
4. BoF としての特徴量抽出
   以上で得られる 100 クラスをビンとして，各小袖屏風画像中に含まれるセグメントを単位とするヒストグラムをつくることができる. このようにしてつくられる特徴は，BoF(Bag of Features)と呼ばれている. BoF 同士が近く, ヒストグラムの頻度をベクトル化した場合の内積が大きい画像同士は, 共通のモチーフを包含した画像であることが予想される. この共通性により，様々なモチーフの組み合わせから得られる特徴が分類できる.
5. Tf-Idf による重み付け
   4 で得られる BoF を構成しているビンの中には, 小袖屏風の背景である金箔部分や衣桁などの重要でない特徴も存在する. また，多くの小袖に共通して含まれる特徴的ではないセグメントは, 冗長な成分となる. そこで，各小袖屏風画像の中で程度頻出するかにより, モチーフを特徴づけるクラスかどうかを判定する. そのために，自然言語で使われる Tf-Idf(Term Frequency -Inverse Document Frequency)を用いる. Tf-Idf は式(3)のような重み付けによって求められる.

$$w_{ij} = \frac{tf_{ij}}{\sum_i tf_{ij}} \log \frac{N}{df_i} \quad \cdots\cdots\cdots (3)$$

ここで　w_{ij}は，j番目の小袖画像におけるi番クラスの重みである. tf_{ij}はj番小袖画像におけるi番クラスセグメントの出現回数，df_iはi番クラスが存在する屏風の数，Nは屏風の総数である. この重みをヒストグラムの頻度に乗ずることにより, 当該画像に特徴的なクラスを強調する.

図 6 に以上の処理の流れを示す.

ⓒ 2015 Information Processing Society of Japan

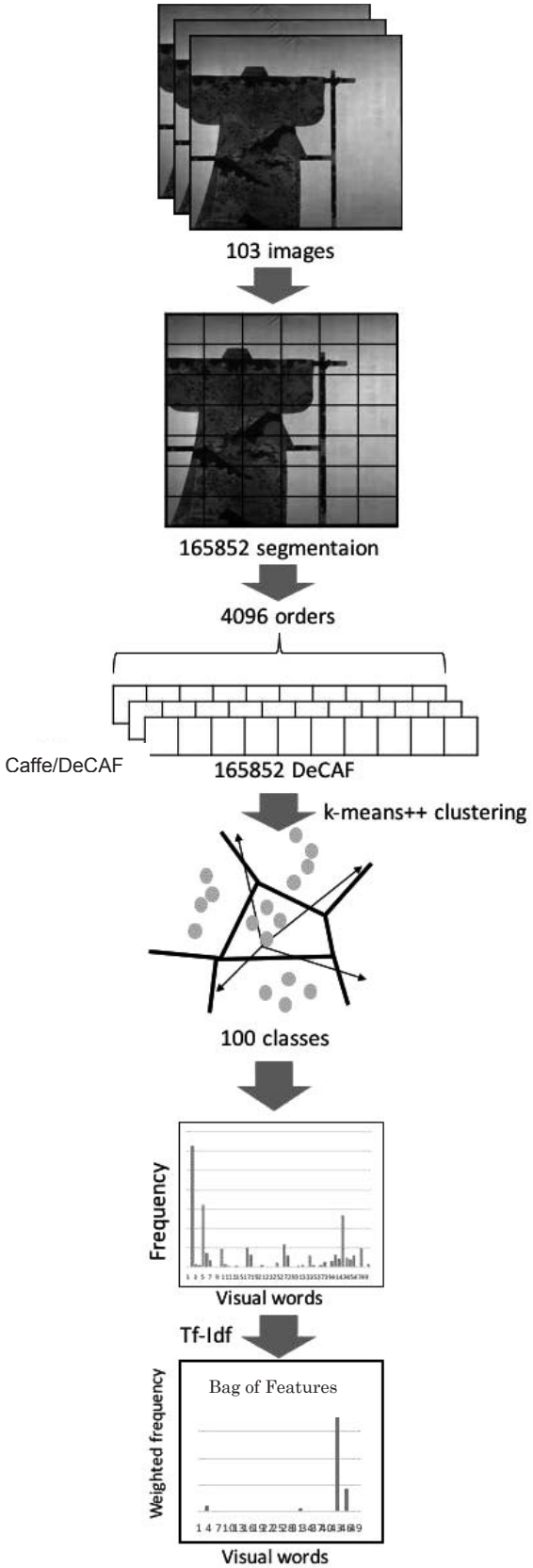

図 6 Caffe/DeCAF による小袖屏風の特徴分析

5．実験および結果

5.1 小袖屏風画像の類似性評価

小袖屏風画像ごとに BoF を求め，その相関に基づき画像間の類似性をネットワークで表現した．具体的には，閾値以上の相関値を持つ画像同士が類似しているとみなす．それぞれの画像のうち，もっとも BoF が近いもの同士が直接リンクされている．

図 7 に，相関の閾値に対して分類されたクラスタ数を示す．閾値が小さい場合は大きなクラスタが形成され，総クラスタ数は小さい．また，閾値が大きくなるとクラスタが形成されずやはり総クラスタ数は小さい．

本実験では閾値が 0.5～0.65 の範囲がもっとも特徴的なクラスタが形成されやすいという仮定をもうける．

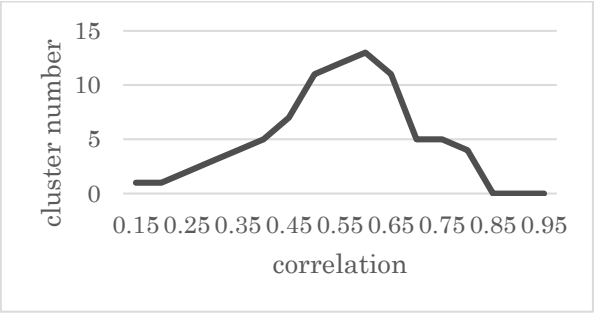

図 7 相関の閾値と得られるクラスタ数の関係

図 8 に，102 隻の画像のうち，相関 0.55 以上の類似度を有する画像同士を関連づけたネットワークを示す．2 つ以上の画像からなるクラスタは 8 つ得られた．

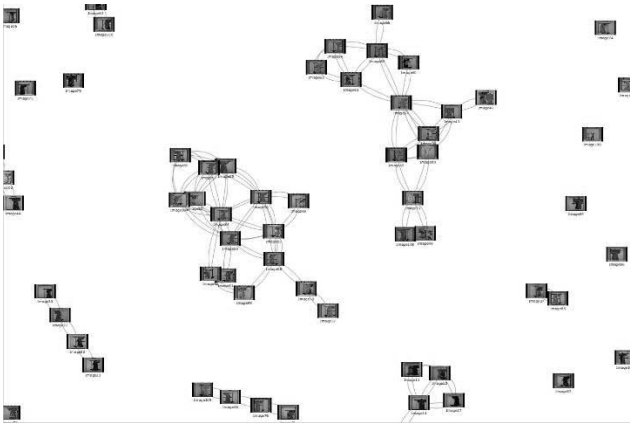

図 8　小袖屏風の類似性クラスタネットワーク

ⓒ 2015 Information Processing Society of Japan

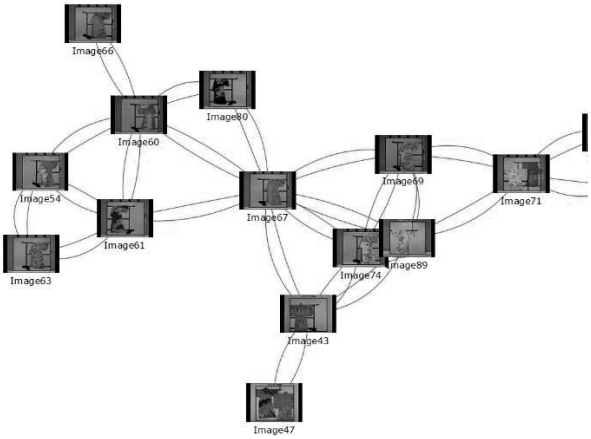

図 9　クラスタ 1（菊のモチーフが主題）

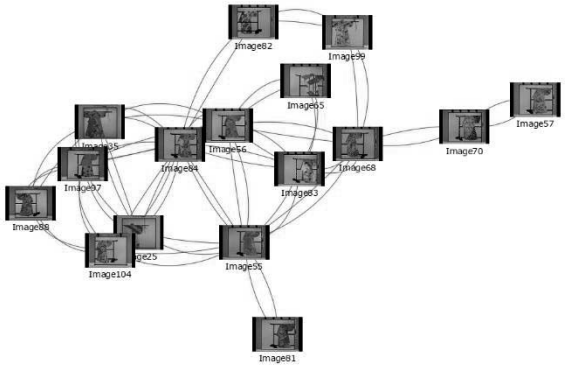

図 10　クラスタ 2（梅・桜のモチーフが主題）

図 9,10 に，ネットワーク中に現れる代表的なクラスタ部分を示す．それぞれ，菊文を主題とする画像のクラスタ，桜文・梅文を主題とする画像のクラスタとなっていた．

つぎに，それぞれの画像の BoF と，特徴的なクラスに対応するセグメントの比較を行った．図 11, 12 に，２つのクラスタの BoF ヒストグラムとそこに含まれるセグメントを示す．これらのヒストグラムは，クラスの出現画像数に対して降順にソートされている

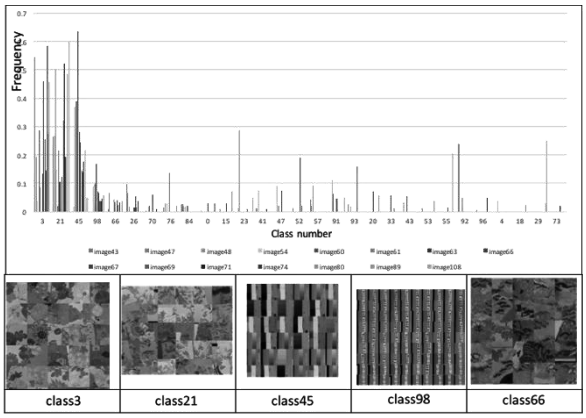

図 11 クラスタ 1 の BoF とセグメント

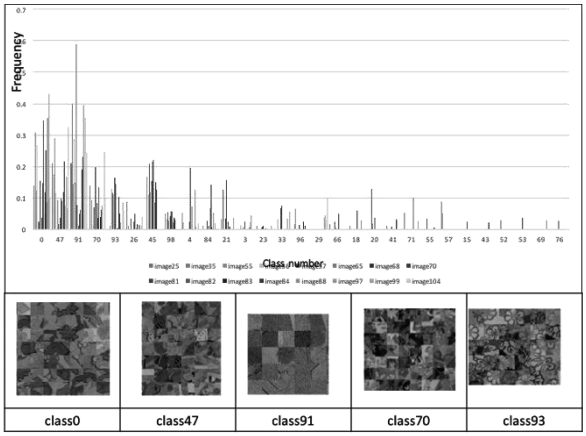

図 12　クラスタ 2 の Bof とセグメント

本結果から，それぞれの BoF の主たるクラスは, およそ 5～10 程度のクラスのまとまりが主要な特徴になりうることが明らかとなった.

5.2 小袖屏風の主題との比較

クラスタに含まれる小袖屏風画像とその主題の対応を図 13，14 に示す．モチーフこれらの画像の主題は，あらかじめ専門家によってタグ付けされたデータである．クラスタ 1 では 15 枚中 11 枚において菊のモチーフのタグがつけられていた．また，クラスタ 2 では 16 枚中 7 枚で，桜・もしくは梅のタグがつけられていた．

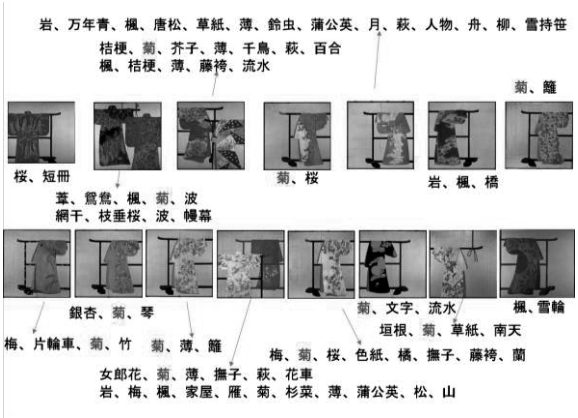

図 13　　クラスタ 1 のメタデータとの対応

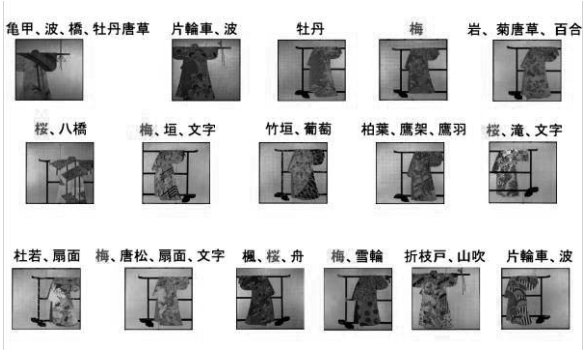

図 14　クラスタ 2 のメタデータとの対応

ⓒ 2015 Information Processing Society of Japan

６．考察

クラスタ 1 に特徴的な Class 3 は，一輪の菊の花弁を正面からみたときのクラスになっていた．また，2 番目の特徴である Class21 は枝を含む横からみた菊の花弁など，一重菊のような小紋に該当するクラスになっていた．Class66 は，松文に相当するクラスであった．

クラスタ 2 に特徴的な Class0 は，特定のモチーフに対応はせず，鹿子絞りの文様のパターンの類似性でまとめられている．同様の傾向は Class91 も該当する．Class47 はセグメントを斜めに分断する模様のクラスであり，扇子や花弁の一部が該当している．Class93 は桜や梅の花弁に該当している．

一方，Class45 や Class98 のように，撮影時のカラーチャートやラベルデータなど，明らかに屏風画像とは無関係なセグメントも特徴として用いられている．本来これらのセグメントは Tf-Idf の処理によって除外されるべきものであるが，k-means によるクラスタリングの際に，複数クラスに分かれている場合は，除去されない場合もある．

また，Class70 のように一見共通性が認められないクラスも特徴として用いられている．

本分析は，モチーフを軸とした小袖屏風画像の特徴を分類している．特定のモチーフを含むことが条件ではなく，モチーフの集合とそのヒストグラムのパターンの類似性を評価しているため，必ずしも共通モチーフの存在が重要ではないが，本結果から，主題のモチーフにそったクラスタを自動抽出できていることを確認することができた．

７．あとがき

深層学習を用いたモチーフ分析にもとづく，小袖屏風画像の特徴分析を行った．

本手法では，深層学習器の中間層に現れる特徴ベクトルをもとに，モチーフの特徴を分類し，その集合によって小袖屏風画像の類似性を定量的に評価できることを示した．

得られた類似のセグメント中には，関連する主題が認められたが，モチーフのみならず，技法の特徴や，印象として明確に表現されないセグメントも同時に抽出されていた．

今後は，セグメントのクラスタリングの方法を工夫し，目的とするメタデータを効率的に抽出する学習が必要である．また，専門家によって抽出されたモチーフのデータや，他のメタデータとの付け合せによる評価を行う．

謝辞 本研究の一部は，国立歴史民俗博物館共同研究（平成 25 年度～平成 27 年度）「歴史資料デジタルアーカイブデータを用いた知的構造の創生に関する研究－小袖屏風を対象として」[6]の支援による．

参考文献

1) 濱上知樹: 歴史資料デジタルアーカイブデータからの知的構造の抽出 - 小袖屏風を対象として - ，電気学会システム研究会, ST-13-124
2) 国立歴史民俗博物館：野村コレクション (1990)
3) 濱上知樹，澤田和人: 小袖屏風を対象とした知的構造の抽出に関する研究，研究報告人文科学とコンピュータ 2015-CH-107,No6,pp 1-4(2015).
4) Caffe: http://caffe.berkeleyvision.org/
5) Jeff Donahue et.al: DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, In Proceedings of the International Conference on Machine Learning (ICML), Beijing, China, June (2014).
6) 歴史資料デジタルアーカイブデータを用いた知的構造の創生に関する研究 －小袖屏風を対象として
http://www.rekihaku.ac.jp/education_research/research/list/joint/2013/digitalarchive.html

ⓒ 2015 Information Processing Society of Japan