

伝統文様アノテーション自動化のための 自然画像とフラクタル画像による事前学習

鏡川 悠介^{1,a)} 久保山 哲二^{2,b)} 加茂 瑞穂^{3,c)} 前田 英作^{4,d)}

概要：本研究では、伝統文様のデジタルアーカイブ化に資するアノテーション自動化手法について検討した。デジタル化された伊勢型紙約 18,000 枚に描かれた文様（梅、桜、菱など）を手がかりに、型紙の文様（約 300 種）の自動識別を試みた。伊勢型紙は、文様の抽象度が高いだけでなく、同種の文様でも型紙ごとにデザインが大きく異なる。さらに、伊勢型紙はほぼ白黒 2 値で表現されているとともに、自然画像のような仔細なテクスチャを持たない。そのため、自然画像による事前学習モデルを用いたニューラルネットワーク (CNN) による従来手法では十分な性能を得ることが難しい。本研究では、自動生成したフラクタル画像による事前学習モデルを用いたアンサンブル学習によるアノテーションの自動化を試みた。

キーワード：伊勢型紙, デジタルアーカイブ, フラクタル, 画像分類

Automated Annotation of Stencil Images using Pre-trained models with Natural and Fractal Images

YUSUKE KYOKAWA^{1,a)} TETSUJI KUBOYAMA^{2,b)} MIZUHO KAMO^{3,c)} EISAKU MAEDA^{4,d)}

Abstract: In this paper, we propose an automated annotation method for the digital archiving of Ise-Katagami, Japanese traditional stencils. We tried to automatically classify the object types depicted in the stencils (about 300 types) based on the patterns (plum blossoms, cherry blossoms, water chestnuts, etc.) for about 18,000 digitized stencil images. The designs in the stencils are not only highly abstract but also highly diverse even within the same class of objects. Moreover, the stencils are almost monochrome, binary black-and-white images, and do not have rich texture information such as in natural images. Therefore, it is difficult to achieve sufficient performance with conventional methods using pre-trained neural network models on natural images. Thus, we propose an improved ensemble method using both fractal images and natural images for pre-trained models.

Keywords: Ise-Katagami, Digital Archives, Fractal Images, Image Classification

1. はじめに

伊勢型紙は着物などの布製品へ文様を染色するための道具である。三重県鈴鹿市を主産地とし、そのデザイン性や彫刻技術が認められ日本の重要無形文化財に指定されてい

る。型紙は消耗品であり、さらに染色の回数を重ねるに従って傷んで使えなくなってしまう。

型紙は資料に直接残された文字情報が極端に少ないため、制作当時何を表現しようとしていたのかを正確にたどることは難しい。しかし、型紙はデザインの宝庫であり、デザインを通じて服飾文化や文様に対する意識、あるいは教養をも明らかにする可能性を秘めた資料である。型紙の研究や有効利用のためには、型紙に何が表現されているのかをメタデータとして蓄積していくことが必要不可欠となる。一方、型紙は数千・数万枚単位で保管されるため、デザインに含まれる文様や技法の情報等を人手により一貫性

¹ 東京電機大学大学院情報環境学研究所

² 学習院大学計算機センター

³ 武庫川女子大学付属総合ミュージアム

⁴ 東京電機大学システムデザイン工学部

a) 20jkm10@ms.dendai.ac.jp

b) ori-jinmoncom2021@tk.cc.gakushuin.ac.jp

c) mizuho.k83@gmail.com

d) maeda.e@mail.dendai.ac.jp

のある基準に基づいて記録していくには膨大な時間と労力がかかってしまう。そのため、これらの作業の自動化ができれば、型紙の保存、研究、利用促進に大きく寄与することが期待できる。

こうした背景から優れたデザインの宝庫である伊勢型紙のデジタルアーカイブ化が進められている。中でも株式会社キョーテックの型紙コレクションをもとに、立命館大学アートリサーチセンターにおいて加茂らにより作成されたデジタルアーカイブ [1][2] には約 18,000 件のデータが保管され、手作業によるアノテーションが行われている。アノテーションには型紙中に出現する図柄の分類ラベルが、図 1 のような階層構造に従って付与されている。具体的には 1 つの型紙内に現れる文様に応じて菊、桜、緋などの文様のラベルが複数付与されている。また、型紙の制作に用いられている技法等もアノテーションとして含まれている。本研究は型紙の文様データのアノテーション自動化のための高精度な手法の開発を目的としており、そのために必要となる、型紙中に含まれる図柄の分類手法を提案する。

画像認識の研究分野においては畳み込みニューラルネットワーク (CNN) を用いた学習モデルが自然画像における物体認識に対して高い性能を誇っている。2012 年の ILSVRC (ImageNet Large Scale Visual Recognition Challenge) [3] での CNN に基づく手法の成功を皮切りに第三次 AI ブームが始まった。大規模なラベル付き自然画像データセットである ImageNet [4] は 2012 年の AlexNet [5] や VGG16 [6] や ResNet [7] などの学習モデルを誕生させるきっかけとなった。一般的に、画像認識では訓練用のデータが大量にある自然画像を用いて事前学習したモデルを初期モデルとして用いる。この初期モデルに特定のタスクに特化したデータを適用して目的に即した学習モデルを得ることを fine-tuning と呼び、物体認識のみならずあらゆるタスクにおいても利用されている。本研究でも同様に、型紙の文様を認識するために自然画像による事前学習を用いた手法が考えられる。しかし、型紙の文様認識に際しては、次の 3 つの課題がある。

- (1) 事前学習に用いられる大規模自然画像データセットの ImageNet は非商用利用や教育目的に関してのみその利用が許可されており、型紙の文様認識を一般利用する際の障壁となる。
- (2) CNN は形状よりもテキストチャに基づいて認識をするテキストチャバイアスの傾向があることが知られており、テキストチャ情報の少ない型紙の認識では十分な性能が得られない可能性がある。
- (3) データ数に対して文様表現が多様であり、学習が困難である。図 2 の例に示すように同じ蝶というラベルが付与されていてもその描かれ方は様々である。

本研究ではこれらの課題を解決するために、ImageNet の代わりに自動生成されたフラクタル画像による事前学習モ

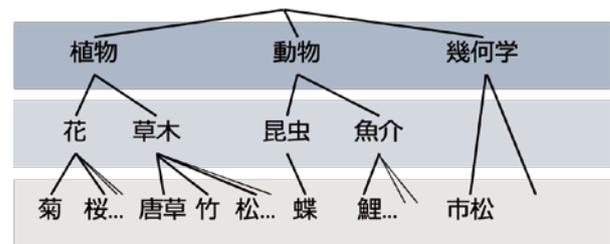


図 1: ラベル構造
Fig. 1 label tree

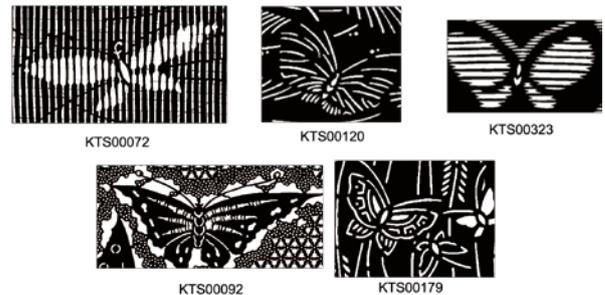


図 2: 伊勢型紙の表現の多様性
Fig. 2 A variety of IseGatagami

デル [8] を用い、型紙の認識を試みた。フラクタル画像データセットは ImageNet の肖像権の侵害やラベルの偏りなどの諸課題を解決する人工生成データセットである。フラクタル画像データセットは大量のフラクタル図形と、図形に対応するラベルを自動生成した訓練用画像を用いる。これにより、いくつかの fine-tuning タスクにおいて ImageNet による事前学習モデルによる識別性能を上回った。フラクタル画像を用いることで大量のアノテーションデータが不足している型紙において、多様な文様表現のパターンが生成可能であり、型紙の抽象的なデザインを識別するための特徴表現の獲得が可能になると期待できる。本研究では人手によるアノテーションが不要なフラクタル画像による複数の事前学習モデルをアンサンブル学習による組み合わせで ImageNet による事前学習モデルと同等の認識性能を達成した。

2. 既存研究

2.1 大規模データセットによる事前学習

昨今の Deep Learning ブームの火付け役となった大規模自然画像データセットである ImageNet [4] は 2012 年の AlexNet に始まり、VGG16 や ResNet [7] などのあらゆる識別モデルを誕生させるきっかけとなった。現時点での深層学習は大量データを用いた事前学習をベースにモデルを作成し、学習する方法が主流であり、ImageNet は事前学習で最も用いられているデータセットである。本研究においても CNN を用いた伊勢型紙画像を認識するために ImageNet の事前学習を使用した。しかし、豊富な色やテキストチャ情報を持つ自然画像に対して、伊勢型紙は全く異

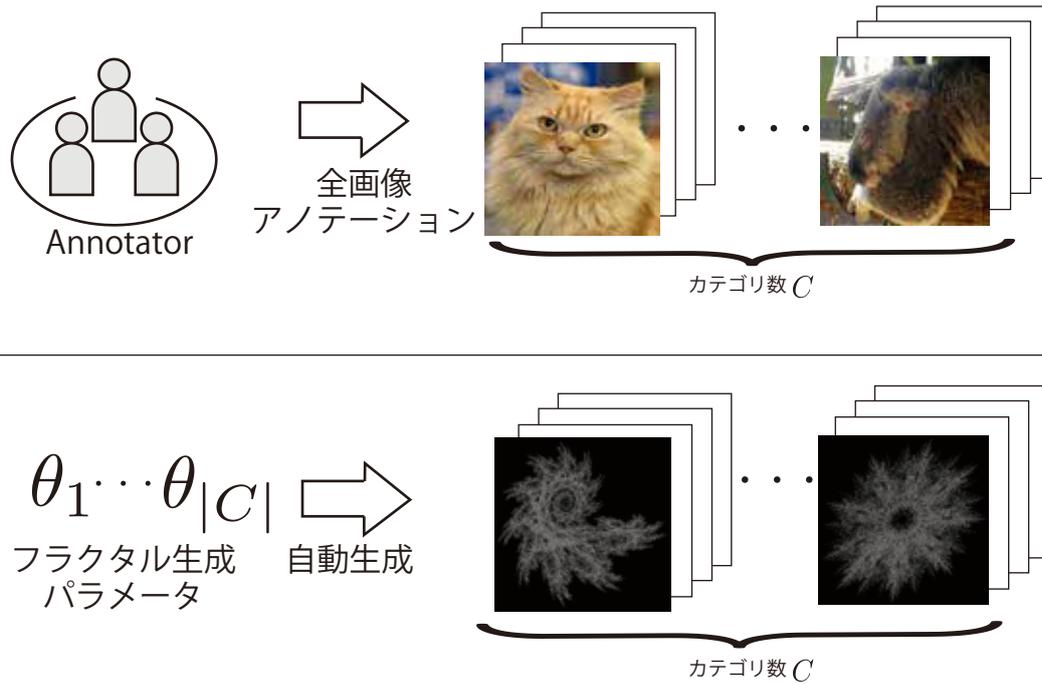


図 3: ImageNet[4] と FractalDB[8] のデータセット構築の比較

Fig. 3

なる性質を持つ画像であるため事前学習モデルが十分に活かされない可能性がある。そこで我々は人工生成データセットである FractalDB [8] を用いて伊勢型紙の抽象的かつ幾何学的なパターンの認識を試みる。フラクタルデータセットは ImageNet の肖像権やアノテーションコストの諸課題を解決するデータセットであり、大量の画像とラベルの対はフラクタル幾何学に基づく生成および画像変形に基づき自動的に生成される。ImageNet と FractalDB のデータセット構築方法について図 3 に示す。ImageNet ではデータセットに含む自然画像を収集し、そのすべての画像に対して人手によってカテゴリが付与されている。片岡らが提案する FractalDB [8] はパラメータによってカテゴリを決定し、カテゴリ内では画像変形に基づき画像が生成される。FractalDB は無数のフラクタル画像をパラメータ調整によって完全自動で生成する。これにより FractalDB は ImageNet のような人手によるアノテーション作業なしで自然物認識のための大規模事前学習データセットの構築を行う。片岡らはフラクタル画像を事前学習データセットとして用い、ImageNet を用いた場合と同程度の正解率を様々な fine-tuning タスクにおいて確認している。

2.2 仔細なテクスチャを持たない認識

伊勢型紙は仔細なテクスチャをもたない。昨今の CNN の学習において重要視されているのはテクスチャであることがわかっている [9][10]。仔細なテクスチャを持たないデータセットを対象にした研究として漫画画像の認識が

ある。坪田ら [11] は ImageNet を事前学習した ResNet を用いて漫画キャラクターの顔画像認識の特徴抽出を行った。ImageNet は漫画画像の認識に一定の成果を挙げているが、実際のところ Manga109 [12] に見るような大規模データセットが整備されているため、ターゲットとするドメインの fine-tuning 用データセットが豊富に存在する。また、漫画は作品内でキャラクターを意図的に描き分けられているため、同一クラス内の類似性は高い。型紙はその点同一クラス内でも描かれている絵柄の表現が多様している。本研究はテクスチャ情報が少なく、抽象度の高い文様を持つ型紙の認識を目標とする。

3. 伊勢型紙データセット

3.1 伊勢型紙デジタルアーカイブの意義

本研究で対象とする型紙とは、柿渋を引いた和紙に様々な文様を彫刻し、小紋や浴衣などの染色に使用するものである。昨今は海外の博物館や美術館に所蔵される型紙の調査が続々と開始された。近年の型紙研究では地方での型紙の流通や所蔵調査について研究が進められてきた。しかし、型紙は数万枚という単位で所蔵されるため、基礎的な調査のためには膨大な時間と労力が要求される。そのため、多くの機関が所蔵調査の段階でとどまっている。型紙は材質は紙であるが、かさねれば重く、大量の型紙を移動することは容易ではない。そこで、型紙のデザインを研究するツールとして画像データの閲覧と検索が可能なデータベース構築が求められる。データベース構築の目的は 2 点ある。

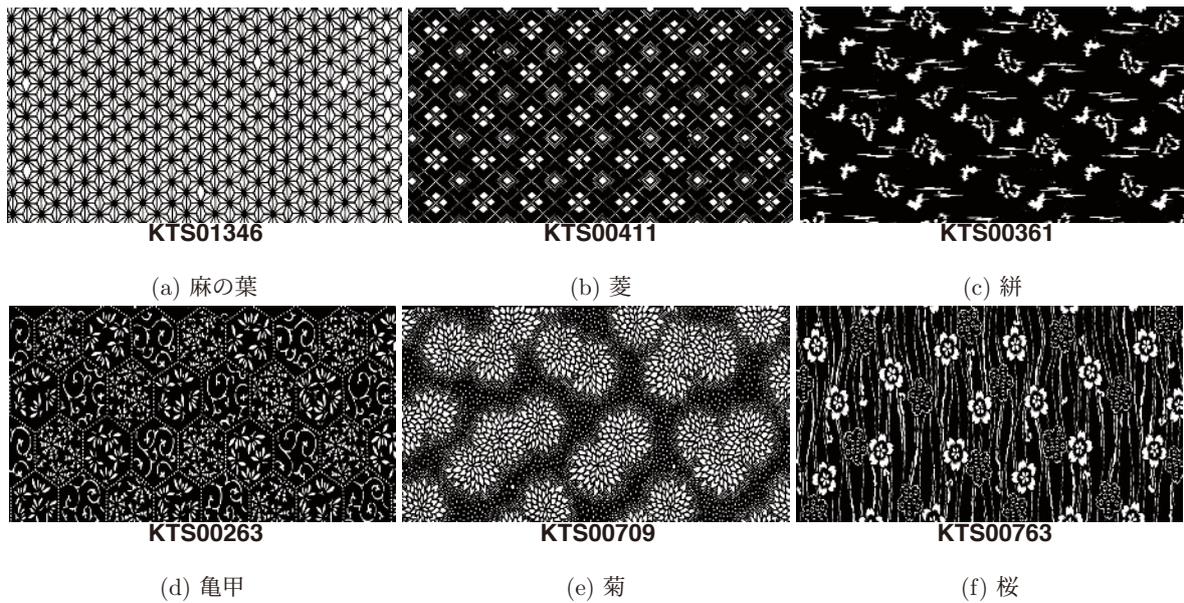


図 4: 型紙文様 6 クラス (麻の葉, 菱, 緋, 亀甲, 菊, 桜) の例

1 点目は型紙をコレクション毎にデジタルアーカイブすることで、コレクションの比較や資料の特質を明らかにする研究を進展させるためである。2 点目はデータベース構築による対象資料のジャンルを超えた研究の進展である。たとえば、型紙と浮世絵に描かれた着物の比較研究が想定される。このように、伊勢型紙のデジタルアーカイブ化は日本の伝統文化の継承のみならず型紙を美術・デザインの観点から分析を可能にすることで美術史研究にも貢献することが期待できる。また、昨今のデータサイエンスと人文科学ドメインの融合 [13][14] に見られるように情報学的視点から伝統文様を解析することでこれまで見られなかった知見の獲得や伊勢型紙そのものの認知向上や普及に寄与するものだと考える。

3.2 デジタルアーカイブのためのラベル付

株式会社キョーテックの型紙コレクションをもとに、加茂らにより作成されたデジタルアーカイブ [2] には約 18,000 件のデータが保管され、手作業によるアノテーションが行われている。コレクション中の型紙には複数の絵柄が描かれており、その絵柄や彫刻技法に対してラベルが付与されている。ラベル構造は大中小のツリー構造であり、大分類には動物、幾何学など。中分類には昆虫、小分類には蝶などのラベルが付与されている。ラベルデータのツリー構造を図 1 に示す。

3.3 データの特徴

本研究で取り扱う型紙文様を学習・分類する上での課題は以下の 3 点である。まず、白黒 2 値で自然画像のようなテキストチャが少なくことである。次に、作者によって創意工夫が込められており、同じ蝶というラベルであっても

描かれ方が異なる。表現の多様性を持つ例を図 2 に示す。図 2 に示すように同じ蝶というラベルが付与されていてもその描かれ方は様々である。最後に多様性のあるラベルデータを学習する際にデータ数が微小である。

3.4 学習用データセットの作成

本研究では、小分類の代表的な 6 つの文様 (麻の葉, 菱, 緋, 亀甲, 菊, 桜) を認識することを検討した。小分類は大分類中分類とは異なり下位概念を含まないため、型紙に含まれる図柄の種類に限られるからである。6 つの文様を認識する学習モデルを構築するためのデータセット作成のために我々は以下のことを行った。

はじめに、型紙に含まれる絵柄領域の抽出を行った。型紙は和紙全体に絵柄を彫刻することはなく、和紙の中心領域に絵柄が描かれている。図柄以外の余白領域を学習データに含めないために、余白領域のみを抽出し、絵柄領域のみの画像を作成した。これを株式会社キョーテックの型紙コレクション 18,000 件全てに対して行った。

次に、小分類に属する代表的な 6 つの文様 (麻の葉, 菱, 緋, 亀甲, 菊, 桜) から約 100 枚の型紙を抽出した。今回取り扱った伊勢型紙 6 クラスのサンプル画像を図 4 に示す。1 つの型紙には文様クラスを表す複数の絵柄が含まれている。

最後に、各型紙画像に対し 512×512px のパッチ領域をランダムに 500 枚切り出し学習データセットを構築した。型紙画像は 1 辺が 1000px を超える高解像度画像であり、そのままでは ImageNet の事前学習モデルを利用し、fine-tuning をすることはできない。このような場合、リサイズ処理によって画像を縮小化することが考えられるが、型紙に描かれる図柄の大きさは型紙によって異なる。その

ため単純なリサイズ処理を施すと細かな点や特徴的な絵柄自体も縮小化されて、情報の欠落が起きてしまう。そこで、1枚の型紙画像中からランダムにパッチ画像を切り出す処理を施した。1枚の型紙画像中には学習対象の様相が複数描かれているため、ランダムに切り出すことで1クラス100枚の画像データの拡張になることを期待した。

4. 手法

学習には大規模データセットによる事前学習モデルを使用する。取り扱う事前学習データセットは以下の3つである。(1) ImageNet(**IN**): 1000クラス各1000枚の100万枚の自然画像。(2) Fractal-Gray(**FG**): 文献[8]のパラメータをもとに生成した1000クラス各クラス1000枚のグレースケールフラクタル画像。(3) Fractal-Binary(**FB**): FGのデータを2値化によってモノクロにした画像。また、事前学習の有無を検証するために初期モデルの重みをランダムに設定したRandom(**RN**)も用意した。モデルの学習と推論の概略図を5に示す。学習をする際に、型紙画像からランダムに切り出したパッチ領域に対してクラスラベルを付与しモデルの学習をした。推論をする際には学習のときと同様に型紙を切り出すが、各パッチ領域に対する予測の算術平均をとることで型紙画像の予測ラベルとする。学習モデルにはResNet18[7]を用いた。fine-tuningした際の各シングルモデルとそれらをアンサンブルしたモデルでIN単体との正解率とF1スコアを比較する。

使用言語はPython, モデルを作成する際のフレームワークにPyTorch [15]を使用し, 学習する際のGPUマシンにはNVIDIA A100(40GB)×4を使用した。

5. 実験

本章では型紙画像に対するfine-tuning実験について述べる。

5.1 実験内容

伊勢型紙に対する事前学習モデルの有用性の検証を行った。6つの文様画像(麻の葉, 菱, 緋, 亀甲, 菊, 桜)各100枚の画像を用意し, そのうちの60%を訓練画像, 60%を検証用画像, 20%をテスト画像とした。テスト画像に対する予測結果から型紙の認識タスクにおける自然画像とフラクタル画像の分類精度とその傾向の分析を行った。

比較項目にはIN, FG, FB, RNのシングルモデル。また, これらシングルモデルの予測結果をアンサンブルしたモデルを用意し, IN単体のモデルとの精度を比較する。アンサンブル手法には[16]を参考にし, 各モデルの全結合層の出力を平均する場合(**FC**)と, 全結合層の後にソフトマックス関数を通して得られた出力を平均する場合(**SM**)で分けた。平均には算術平均(**mean**)と幾何平均(**gmean**)の2種類を用いた。

5.2 学習条件

本実験では伊勢型紙の特徴を学習するCNNモデルとしてResNet18を使用した。片岡らの論文ではResNet50を使用した際にImageNetと同精度の認識性能を誇ったが本研究ではfine-tuningのデータ量を考慮してResNet18を使用した。ResNet18を3つのデータセットIN, FG, FBとランダムイニシャライズであるRNの計4つの手法で事前学習し, 伊勢型紙データセットでfine-tuningする際のモデルの初期値とする。

fine-tuningを行う際には小分類6クラスのデータセットからランダムに512×512pxのパッチ画像を切り出し, 224×224pxにリサイズし学習した。パッチ領域に対して学習を行い, 推論時には複数枚切り出した画像をアンサンブルすることで画像サイズの大きい型紙に対しても予測を行う。推論方式の概略図を図5に示す。最適化手法はAdam [17]を用い, 学習率は0.003, epochは100とした。

5.3 結果

表1よりINが最も高い性能となった。シングルモデルにおいてはINを事前学習に用いた場合が最も高い正解率となった。INは麻の葉, 亀甲で最も高いF1スコアを記録した。アンサンブルモデルではRN以外の3つのモデルを全結合層の出力直後に算術平均をしたものが78%と最も高い正解率となった。IN以外の3つのモデルのアンサンブルではソフトマックス関数のあとに幾何平均をしたもの(SM/gmean)が74.6%とIN単体の正解率を上回る結果となった。

5.4 考察

表1の結果から大規模事前学習データセットを単体で用いるよりもそれらを組み合わせることで精度向上を確認した。INを用いずに, フラクタル画像の事前学習モデルを組み合わせたモデルを用いることでIN単体と同等あるいはそれ以上の精度向上が期待できる。しかし, INを含むアンサンブルモデルとINを含まないアンサンブルモデルでは予測ラベルのアンサンブル位置と平均のとり方によって精度の上がり方に違いが見られた。しかし, アンサンブル手法の良否を決定付ける結果は本実験からは得られなかった。

ImageNetとFractalDBの違いを評価するために, Grad-CAM [18]による注視領域の可視化を行った。Grad-CAMによる可視化結果を図6に示す。CNNが予測のために注視した領域がヒートマップで示されている。画像中の赤い領域はCNNがラベルを予測する際により注視した領域である。菊ラベルに対しては, INの方がより菊の絵柄に注視しており(a), FGは周辺の小さなパターンに集中していることがわかる(b)。表1のクラスごとのF1スコアも参照するとINが絵柄に対する認識性能が高いことがわかる。菱ラベルに対して, INでは図中の1箇所のみ菱形を注視

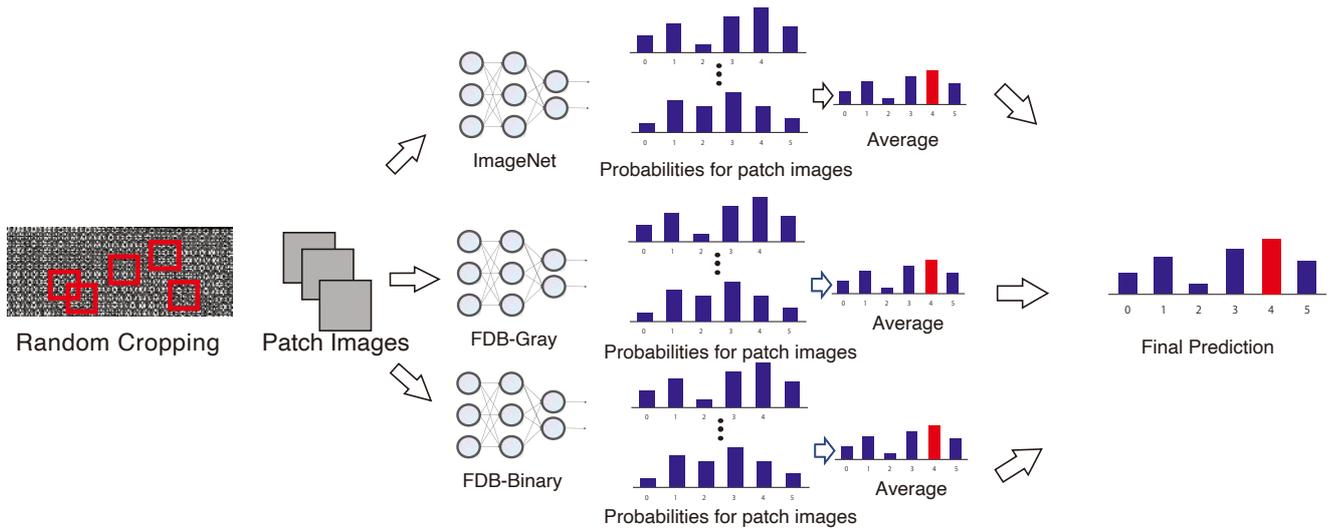


図 5: 型紙文様認識のためのアンサンブル手法

Fig. 5 A ensemble method

表 1: シングルモデルとアンサンブルモデルのクラスごとの F1 スコアと正解率

Table 1 F1 scores and accuracies for each class of single and ensemble models

Method	Fusion/Avg	F1-score/class						Accuracy (%)
		麻の葉	菱	緋	亀甲	菊	桜	
ImageNet(IN)	N/A	0.90	0.55	0.85	0.71	0.71	0.64	72.8
Fractal-Gray(FG)	N/A	0.74	0.67	0.76	0.59	0.53	0.60	65.2
Fractal-Binary(FB)	N/A	0.67	0.60	0.71	0.43	0.48	0.50	56.7
Random(RN)	N/A	0.55	0.53	0.80	0.48	0.65	0.57	59.3
FG+FB+RN	SM/mean	0.76	0.72	0.82	0.60	0.63	0.72	71.2
FG+FB+RN	SM/gmean	0.75	0.71	0.87	0.70	0.65	0.78	74.6
FG+FB+RN	FC/mean	0.72	0.67	0.87	0.65	0.61	0.78	72.0
IN+FB+RN	SM/mean	0.88	0.67	0.83	0.63	0.77	0.68	74.6
IN+FB+RN	SM/gmean	0.79	0.67	0.95	0.65	0.77	0.76	76.3
IN+FB+RN	FC/mean	0.82	0.69	0.95	0.60	0.75	0.76	76.3
IN+FG+FB	SM/mean	0.76	0.79	0.95	0.65	0.78	0.71	77.1
IN+FG+FB	SM/gmean	0.85	0.73	0.95	0.70	0.65	0.67	75.4
IN+FG+FB	FC/mean	0.84	0.75	0.97	0.65	0.72	0.74	78.0
IN+FG+FB+RN	SM/mean	0.78	0.75	0.88	0.67	0.78	0.75	77.1
IN+FG+FB+RN	SM/gmean	0.79	0.65	0.95	0.65	0.70	0.72	74.6
IN+FG+FB+RN	FC/mean	0.80	0.60	0.95	0.62	0.70	0.72	73.7

していない (c) が, FG では二箇所ともに注視している (d) ことから FG は従来の IN による学習モデルよりも幾何的
形状に対する高い認識能力を持つことがうかがえる. こう
した学習の違いが表 1 に示した FG, FB, RN のアンサン
ブル学習において IN に近い正解率になった要因であると
考える. 今後の研究方針としてはこのような絵柄パターン
に対するフラクタルによ事前学習モデルの認識性能の向上

が求められる.

6. おわりに

本研究では伊勢型紙のデジタルアーカイブに向けた文様
アノテーション自動化のために CNN を用いて型紙画像の
学習・分類を行った. 非商用利用に限られた大規模自然画
像データセットである ImageNet で事前学習した CNN モ

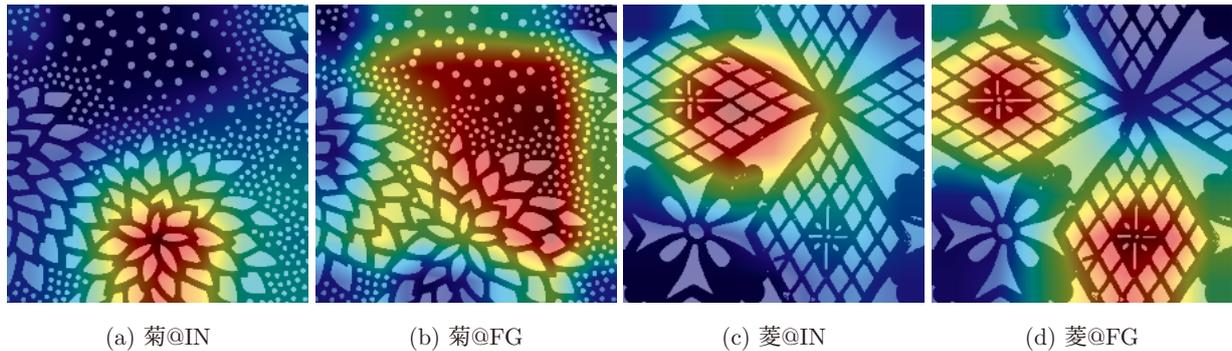


図 6: GradCAM による可視化結果

デルを上回る認識精度をアノテーションを必要としない人工生成パターンであるフラクタル画像を用いて達成した。異なるパターンを得意とするモデルをアンサンブルすることで型紙のような抽象度の高い文様を認識することができる。本研究で用いた手法は伊勢型紙に限らない一般的な意匠データに関して有効であると考えている。今後は今回分類対象とした型紙をより定量的に表すことで、自然画像による事前学習モデルとフラクタル画像による事前学習の学習の傾向の違いを議論することである。具体的には画像のテクスチャ性、繰り返し性、均質性、不規則性を計算する Dai ら [19] の指標を用いて型紙画像を定量的に表すことで可能であると考えている。また、フラクタル画像の事前学習手法に関して工夫の余地があり、Anderson [20] のフラクタル画像の事前学習手法の改善に見るようにフラクタル画像を1画像内に複数配置することで擬似的に型紙文様と同じく複数絵柄を持つようなデータセットが構築可能である。このような複数絵柄を持つ大規模データセットで事前学習したモデルで型紙画像を fine-tuning することで本論文で示した結果を上回るものであると考えている。

謝辞 本研究は「立命館大学アート・リサーチセンター 文部科学省共同利用・共同研究拠点「日本文化資源デジタル・アーカイブ研究拠点」2020年度共同研究の一環として行われた。本研究の遂行に際し、貴重な型紙コレクションの利用を承諾いただいた株式会社キョーテックに謝意を表す。

参考文献

- [1] 加茂瑞穂: ニッポンの型紙図鑑, 青幻舎 (2020).
- [2] 立命館大学アート・リサーチセンター: 型紙データベース (2014). <https://www.dh-jac.net/db1/stencil/index.php>.
- [3] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. S., Berg, A. C. and Li, F.: ImageNet Large Scale Visual Recognition Challenge, *Int. J. Comput. Vis.*, Vol. 115, No. 3, pp. 211–252 (online), DOI: 10.1007/s11263-015-0816-y (2015).
- [4] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255 (2009).
- [5] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States* (Bartlett, P. L., Pereira, F. C. N., Burges, C. J. C., Bottou, L. and Weinberger, K. Q., eds.), pp. 1106–1114 (online), available from <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html> (2012).
- [6] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Bengio, Y. and LeCun, Y., eds.), (online), available from <http://arxiv.org/abs/1409.1556> (2015).
- [7] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition (2015).
- [8] Kataoka, H., Okayasu, K., Matsumoto, A., Yamagata, E., Yamada, R., Inoue, N., Nakamura, A. and Satoh, Y.: Pre-training without Natural Images, *Proceedings of the Asian Conference on Computer Vision* (2020).
- [9] Hermann, K. L., Chen, T. and Kornblith, S.: The origins and prevalence of texture bias in convolutional neural networks (2019).
- [10] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A. and Brendel, W.: ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, *arXiv preprint arXiv:1811.12231* (2018).
- [11] 坪田亘記, 小川 徹, 山崎俊彦, 相澤清晴: キャラクター顔特徴量の個別漫画への適応手法, 人工知能学会全国大会論文集, Vol. JSAI2018 (2018).
- [12] Matsui, Y., Ito, K., Aramaki, Y., Yamasaki, T. and Aizawa, K.: Sketch-based Manga Retrieval using Manga109 Dataset (2015).
- [13] 朝展北本, タリンカラーヌワット, Alex, L., Mikel, B.-I.: くずし字認識のための Kaggle 機械学習コンペティションの経過と成果, じんもんこん 2019 論文集, Vol. 2019, pp. 223–230 (2019).
- [14] Tian, Y., Clanuwat, T., Suzuki, C. and Kitamoto, A.: Ukiyo-e Analysis and Creativity with Attribute and Geometry Annotation, *CoRR*, Vol. abs/2106.02267 (online), available from <https://arxiv.org/abs/2106.02267> (2021).

- [15] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada* (Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. B. and Garnett, R., eds.), pp. 8024-8035 (online), available from (<https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>) (2019).
- [16] Dutt, A., Pellerin, D. and Quénot, G.: Coupled Ensembles of Neural Networks, *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings*, OpenReview.net, (online), available from (<https://openreview.net/forum?id=rylRCUJDG>) (2018).
- [17] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Bengio, Y. and LeCun, Y., eds.), (online), available from (<http://arxiv.org/abs/1412.6980>) (2015).
- [18] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, IEEE Computer Society, pp. 618-626 (online), DOI: 10.1109/ICCV.2017.74 (2017).
- [19] Dai, D., Riemenschneider, H. and Gool, L. V.: The Synthesizability of Texture Examples, *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, IEEE Computer Society, pp. 3027-3034 (online), DOI: 10.1109/CVPR.2014.387 (2014).
- [20] Anderson, C. and Farrell, R.: Improving Fractal Pre-training, *CoRR*, Vol. abs/2110.03091 (online), available from (<https://arxiv.org/abs/2110.03091>) (2021).