

## ディープニューラルネットワークによる画風の特徴抽出 Feature Extraction of Painting Style Using Deep Neural Network

長島秀明\*

Hideaki Nagashima

清水郁子\*

Ikuko Shimizu

### 概要

本論文では、絵画作品に表れる画家それぞれの画風を表す特徴を複数の絵画画像の学習によって抽出することを目的とする。特徴の抽出には大量のデータを用いて学習を行い認識に有効な特徴を自動的に学習することができるディープニューラルネットワーク (DNN) を用いた。画家10名分の絵画画像を用いて特徴抽出を行い、抽出された特徴を用いて認識実験を行った結果、入力画像を分割する等の工夫を施すことで、認識精度が向上することを確認した。

### 1. 序論

画家の画風とは絵画の作風のことであり、絵に表れた画家個人または流派の特色や傾向のことを指す。画風は絵画作品に配置された絵の具の色や筆遣い等で表現されるといわれおり、絵画を写した画像からその情報を得ることができる。

特定の画家が描いた絵画の特徴を学習できれば、その特徴を用いて、他の画家が描いた絵画と識別することが可能であると考えられる。そこで、本研究では、画家が描いた複数の絵画から画家の画風を表す特徴を学習し、その特徴を用いて入力された絵画の画家を認識するシステムを開発し、認識率を調べることによって、学習された特徴量の性能の評価を試みる。

従来の画像認識では、画像から抽出する特徴量をどのように設計するかが重視されてきた。そのため、絵画から画家を認識する従来手法でも、手動で設計した特徴量を用いた研究が行われてきた [1], [2], [3]。

一方で、近年、ディープニューラルネットワーク (DNN) が、様々な分野で高い認識率を達成し注目を集めている。例えば、2012年11月開催の一般物体認識コンテスト Imagenet Large Scale Visual Recognition Challenge (ILSVRC) で、ディープラーニングの方法が従来の方法に大きな差をつけて優勝したことで、研究者たちに大きなインパクトを与えた [4]。また、google の研究 [5] では、YouTube からランダムに切り出した画像を1000万枚使用し、大規模なネットワークを学習することで、人や猫の顔などに選択的に反応するニューロンが自動的に生成されたことが報告され、大きな話題となった。また、音声認識や自然言語処理などにおいても、従来手法を圧倒する性能を示すことが報告されている [6]。高い性能を示す理由のひとつに、

DNN が与えられた課題に応じた適切な特徴量を自動的に学習で得ることができる点が挙げられる。

そこで本研究では、様々な画家が描いた複数の画像を DNN で学習し、画家それぞれの画風を表す特徴を抽出することを試みる。抽出した特徴によって、絵画画像の画家を認識する実験を行い、抽出された特徴量を評価する。

DNN で学習した特徴は、画家の画風を表していると考えられる。DNN の学習には大量の画像が必要であるが、一人の画家が生涯に描いた絵画の枚数は限られている。本研究では、得られた絵画画像のリサイズおよび分割により、データ量を増やすと同時に画像に現れる特徴を様々なスケールで抽出することを試みる。

特徴を学習し、学習した特徴を評価するため、広く知られている西洋絵画の画家たちの絵画画像を収集して識別実験を行う。識別対象とするクラスは、絵画の作者にあたる画家名および流派名とし、テスト用の画像各々が用意したクラスのどれに属するかを識別できるように収集した学習用データセットを用いて DNN を訓練する。本研究では、DNN のうち、画像認識分野でこれまで数々の良い成果をあげている畳み込みニューラルネットワーク (CNN) [7], [8] を用いる。収集した画像を学習用データとテストデータに分割し、認識実験を行った。

### 2. 関連研究

絵画に関する画像処理の分野では、機械学習をベースとした手法が多く用いられており、本研究もこの枠組みに属する。ここでは3つの関連研究について説明する。

Shamir らの研究 [1] では、3つの流派 (印象派、抽象派、超現実主義) に属する9人の画家の作品を分析した。画像特徴はガボールフィルタやエッジ抽出を始めとした様々な特徴を抽出し、フィッシャースコアを使用した評価を行った。画家クラスでの分類精度は77%、3つの流派での分類精度は91%という結果を得た。

Jou らの研究 [2] では、多クラス分類問題の問題への機械学習アプローチが最も潜在的に適しているとしている。HOG と色ヒストグラムを使用した後、Naive Bayes、線形判別分析、ロジスティック回帰、K-means 法及び SVM のようないくつかの多クラス分類手法を適用することで5人の画家分類において65%の分類精度を達成した。

その他にも、絵画分析に役立つ特徴抽出の研究として、Igor らの研究 [3] がある。この論文は絵画画像か

\*東京農工大学工学府, 東京都小金井市中町 2-24-16 Tokyo University of Agri. and Tech. 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan

表 1: 画家とその流派分類

No.	画家名	流派
1	ラファエロ	ルネサンス
2	ミケランジェロ	ルネサンス
3	フェルメール	バロック
4	レンブラント	バロック
5	ドラクロワ	ロマン主義
6	モネ	印象派
7	ルノワール	印象派
8	ゴッホ	ポスト印象派
9	ミュシャ	ポスト印象派
10	スーラ	ポスト印象派

ら筆遣いの方位自動抽出に注目しており、Prevailing Orientation Extraction Technique (POET) と呼ばれるオリエンテーション抽出技術を使用して人間の目では区別できないレベルの筆遣い抽出に成功している。この自動抽出は絵画の筆遣い空間分布の自動抽出を支援することができる。

### 3. 提案手法

#### 3.1. データセット作成

実験では、著作権適用期限外の複数の画家の絵画画像を用いた。油彩を中心として多くの作品を残した有名画家たちの絵画画像を、一人当たり100枚程度、10人分収集した。10人の画家たちは様々な年代で活躍した人物から幅広い流派やジャンルから選択した。画家と流派の分類を表1に示す。

なお、同じ流派に属する画家は画風が比較的近いとされている。実験で画像を収集した10人の画家は5つの流派に属している。

#### 3.2. データセット加工

DNNの学習には一般に数万枚といった多量の学習データが必要となるといわれており、一人の画家当たりの画像枚数が100枚では学習が不十分になってしまう可能性が高い。また、画像のサイズも様々で学習が困難である。

そこで本研究では、学習用データ不足の問題解決と認識精度の向上を目的として2つの工夫をする。

1つ目は、画像の分割である。用意した絵画画像を $32 \times 32$ ピクセルのサイズへと分割し(分割した $32 \times 32$ の画像をパッチ画像と呼ぶことにする。), 合計50万枚程度、画家一人当たり5万枚程度の学習用データセットとする。このように画像を分割することで、学習データを増やすこと、さらにより細かな筆遣いを学習することが可能になる。

2つ目は、画像のリサイズである。収集した絵画画像の多くは実際の絵画をスキャナーもしくは高画質カ

メラで取り込み、そのサイズをさらに縮小したものである。これらをそのまま分割すると、画像の大きさが不揃いであり、さらに、局所的な特徴しか得ることができない。これらの問題点に対応するため、元の絵画画像を $2^n$ 分の1のサイズに縮小してから分割を行う。この処理を行うことで、実際の絵画の大きさと収集した画像のサイズの比が一定でない問題の影響を軽減することができる。さらに、小さな領域だけでなく全体の構図を含めた特徴の学習が期待できる。なお、もとの画像の大きさにもよるが、リサイズによりパッチ画像数は数10%増加する。

#### 3.3. 畳み込みニューラルネットワークによる特徴の学習

本研究では、様々な構造のDNNの中でもで画像の学習に適しているとされるCNN[7], [8]による学習及びテストを行う。CNNは、畳み込み層とプーリング層の2つを交互に繰り返す構造をとる。畳み込み層では入力画像にフィルタを畳み込む。プーリング層ではサブサンプリングを行う。2層を交互に繰り返すことで、フィルタ出力層とプーリング層の2層の出力が、次のフィルタ出力層の入力となる形で積み重ねる。プーリング層での処理はユニット数を減少させるので、上位層ほどユニット数が少なく、最終的に出力層は分類するクラス数分のユニットを持つ。CNNは教師あり学習が前提となっており、SGDを用いた誤差逆伝播法によるパラメータ修正を行うことでネットワークを学習する。

CNNの実装は、EBLearn<sup>†</sup>と呼ばれるライブラリを使用した。EBLearnは、主にCNNによる学習・推論を行うためのライブラリで、C++で書かれている。ネットワークの構成を記述した設定ファイルとデータセットを与えることで、EBLearnが自動的にこれらのデータから各パラメータを学習し、推論結果を返す。

本研究で採用したネットワーク構造は、文字認識や画像認識で高い性能を示しているLenet-5<sup>‡</sup>を参考にした。Lenet-5は6層構造で、 $32 \times 32$ ピクセルのパッチ画像を入力とする。詳細は1層目から、 $5 \times 5$ のフィルタを6個持つ畳み込み層、2層目は $2 \times 2$ 画素領域にサブサンプリングを行うプーリング層、3層目は $6 \times 6$ のフィルタを16個持つ畳み込み層、4層目は $2 \times 2$ の画素領域にサブサンプリングを行うプーリング層、5層目は $6 \times 6$ のフィルタを120個持つ畳み込み層、最終層は全結合の完全接続層である。Lenet-5のネットワーク構造を視覚化したものを図1に示す。

#### 3.4. 学習した特徴を用いた絵画画像の画家認識

本研究では、学習した特徴のよさを評価するために、テスト用の画像の画家の認識を行い、認識率を評価する。

そのために、テスト用の画像も分割とリサイズを行う。すなわち、1枚の絵画画像から精製した複数の $32 \times 32$ の大きさのパッチ画像をCNNへ入力する。各パッチ

<sup>†</sup><http://ebllearn.sourceforge.net/>

<sup>‡</sup><http://yann.lecun.com/exdb/lenet/>

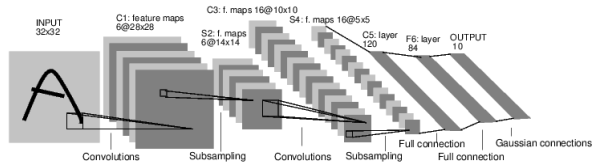


図1: LeNet-5のネットワーク構造 出典: LeNet-5 Demos, <http://yann.lecun.com/exdb/lenet/>, 2014/6

画像に対して認識結果(画家の名前)が出力されるが、本研究では、1枚の画像をリサイズ・分割して得た複数のパッチ画像の認識結果を投票し、投票率の最も高かったクラス(画家)に属していると判定する。

#### 4. 実験結果

提案手法の性能を評価するため、画像を入力し、画家を認識する実験を行った。

##### 4.1. 比較手法

本研究の手法と比較するため、精度評価においてベースラインとして用いられている Naive Bayes ベースの手法を採用した。CNNの学習に用いたものと同様の画像から局所的な輝度勾配である SIFT 特徴量を抽出し、Bag-of-Keypoints による特徴表現に変換後、Naive Bayes を用いてクラス分類を行った。学習では、予め教師データとしてカテゴリごとに分けられている学習画像から SIFT 特徴量を計算し、Bag-of-Keypoints によってベクトル量子化を行う。その後学習と同様にクエリ画像から SIFT 特徴量計算し、ベクトル量子化を経て Naive Bayes によって最も類似度の高いカテゴリに分類する。

##### 4.2. 評価方法

EBLearn での学習によって得られるネットワークにはサイズの上限があるため、学習データを数グループに分けて学習とテストを行った。グループ1は画家別にクラスを分けてあり、4人の画家を含む。グループ2は画家の流派毎にクラス分類してあり、1クラスにつき2人の画家を含んでいる。表2、表3にグループ1、グループ2の分類を示す。グループの分類について、まずグループ1は、特徴がわかりやすい4名の画家(レンブラント、モネ、ゴッホ、ミュシャ)を選択した。グループ2は、4つの流派(ルネサンス、バロック、印象派、ポスト印象派)に対し、各2名の画家のデータを用いた。

用意した2つのデータセットを用いて評価実験を行った。まず実験データセットのグループ1とグループ2について CNN での認識と、比較手法(SIFT+Bag of keypoints と Naive Bayes)での認識を行った。次に、学習に用いるデータセットの枚数を100%、90%、80%、70%、50%、30%、10%に減らし、それぞれの CNN による認識精度を比較した。さらに、学習に用いる画

表2: グループ1

No.	画家名	分類
1	レンブラント	バロック
2	モネ	印象派
3	ゴッホ	ポスト印象派
4	ミュシャ	ポスト印象派

表3: グループ2

No.	画家名	分類
1	ミケランジェロ, ラファエロ	ルネサンス
2	フェルメール, レンブラント	バロック
3	モネ, ルノワール	印象派
4	ゴッホ, スーラ	ポスト印象派

像のリサイズを行った場合と、行わなかった場合の認識精度を比較した。

CNN での認識の流れを説明する。まずグループに含まれる画家の絵画から生成したパッチ画像を用いて学習し、次にそこに含まれていないパッチ画像を識別した。さらに、テスト用データセットの内容によって認識結果に偏りが出るので回避するために、学習用データ及びテスト用データをランダムに入れ替え、その他の条件は変えずに複数回テストを行った。

##### 4.3. 実験1 (CNNと既存手法との比較)

まず、グループ1に含まれる画家(レンブラント、モネ、ゴッホ、ミュシャ)をそれぞれ学習用データで学習し、テスト用データによってテストを行った。提案手法による結果を表4および図2に示す。さらに、比較手法との比較結果を図3に示す。表での誤答率は元の絵画作品単位での認識結果であり、20作品での認識を3セット行った平均値である。適合率平均は $32 \times 32$ のパッチ画像の作品毎の識別結果を平均したものであり、適合率標準偏差はそれらの標準偏差を出したものである。図2で示したのは各作品の適合率が高いものから整列して画家別に並べたもので、縦軸が適合率、横軸が作品番号を示している。さらに図3では CNN と比較手法(SIFT+Bag of keypoints と Naive Bayes)を比較しており縦軸がF値、横軸が画家名を示している。

グループ1の中ではモネの誤答率が最も高く、続いてゴッホ、レンブラント、ミュシャとなった。適合率の平均はモネ、ゴッホ共に40%前後であり信頼性はやや曖昧であったが、誤答率が最も低かったミュシャは適合率平均も70%近く信頼性の高い結果になった。グループ1の中ではその他の画家に比べ、ミュシャの画風がより認識しやすいものということが言えるだろう。

表 4: グループ 1 認識結果

画家名	レンブラント	モネ	ゴッホ	ミュシャ
誤答率 (%)	5.3	35.7	15.2	4.7
適合率平均	54.1	39.0	41.7	68.9
適合率標準偏差	16.6	25.6	19.6	15.4

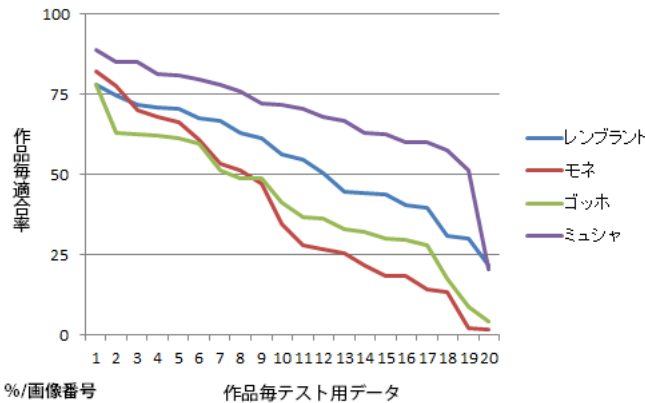


図 2: 高い順に並べ替えたグループ 1 における作品毎の適合率

適合率の標準偏差と 20 枚分の適合率詳細を見ると、レンブラントの認識率の変化は直線的であった。モネは適合率 50% 以上だったところから急激に低下する部分があり、そこからは 25% 前後の判定が多い。ゴッホも同様に段階的に精度が低下し、30~40% 近い判定が多かった。モネ、ゴッホの 2 名は作品の制作年代によって画風の変化が見られ、学習用、テスト用共にデータに偏りがあったためこのような結果になったと考えられる。ミュシャは 1 枚のみ 20% を下回ったがその他は 50% 以上で非常に良い精度で識別することができている。既存手法との比較結果はレンブラント、モネでは既存手法より僅かに性能が下回ったが、ゴッホとミュシャの認識においては大きく上回る性能を示した。

次に、グループ 2 に含まれる画家をそれぞれ学習用データで学習し、テスト用データによってテストを行った。グループ 2 では 1 クラスにつき、同系統の画風または流派の画家を 2 名ずつ割り振ってあり、学習用、テスト用のデータは 2 名で半分ずつ（学習用に一人 2 万枚、テスト用に一人 5 千枚分となる）になるよう調整した。結果を表 5 と図 4、図 5 に示す。

画家の流派毎に分けたグループ 2 の誤答率はルネサンス（ミケランジェロ、ラファエロ）が最も高く、40% を超えた。適合率の詳細を見ると 50% 以下にほぼ一様に分布している。バロック（フェルメール、レンブラン

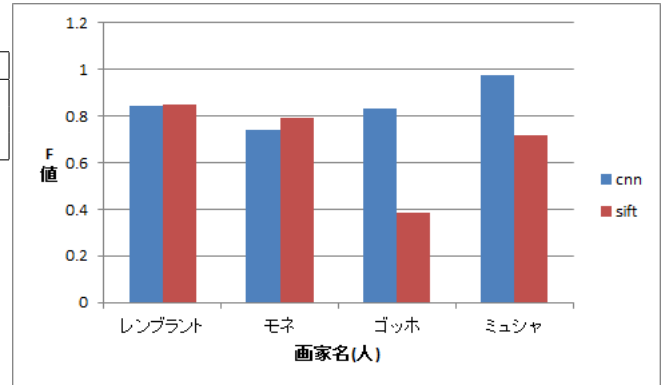


図 3: SIFT+Bok+Naive Bayse との比較-グループ 1

表 5: グループ 2 認識結果

画家名	ルネサンス	バロック	印象派	ポスト印象派
誤答率 (%)	40.5	10.3	25.6	15.8
適合率平均	29.0	47.3	42.7	44.4
適合率標準偏差	12.6	15.4	24.8	27.3

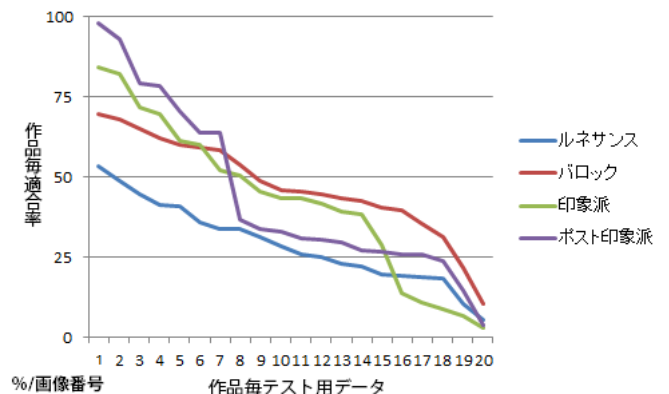


図 4: 高い順に並べ替えたグループ 2 における作品毎の適合率

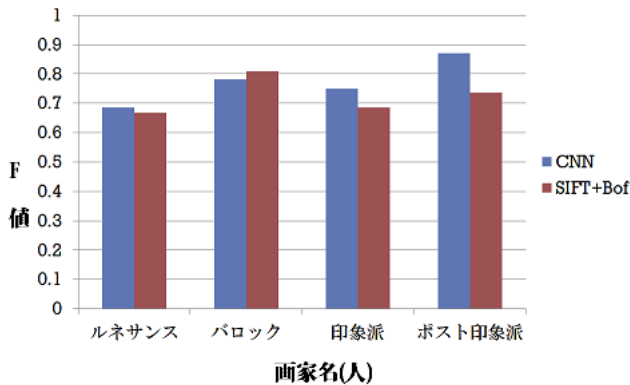


図 5: SIFT+Naive Bayes を使用した手法との比較-グループ 2

ト)は誤答率が低く理想的だが、適合率の平均が50%に及ばなかった。印象派(モネ、ルノワール)は誤答率は25%あるが、適合率平均はバロックに近い40%以上となった。標準偏差の大きさを見れば分かるように、適合率は70%以上のものから一桁台まで広く分布しており、適合率詳細では段階的に変化するグラフから、モネとルノワールで判定が分かれているものと考えられる。ポスト印象派(ゴッホ、スーラ)は印象派よりさらに適合率標準偏差が広がり、適合率詳細の段階的な変化も急激である。グループ2で良い成績を取めたスーラとゴッホの識別難度の違いが顕著に表れたと言える。ゴッホの誤答率が閾値近辺を上下していることから、特徴の抽出はあまり上手くいっていないとも考えられる。既存手法との比較結果はバロックのみ下回ったもののその他ではCNNの方がより良い認識精度を示した。特にポスト印象派では点描を使うスーラの作品が非常に良い精度で認識することができた。なお、流派毎の分類実験で用いたデータセットは、あらかじめ収集した10名分のデータのみを用いたため、実際の流派に属する画家全体の分類ではない。収集した画像は一般的に出回っている作品を使用しているため、画家の作品全てを網羅しているとは限らない。

#### 4.4. 実験2 (学習データを減らした場合)

実験2では、データセットの枚数を100%, 90%, 80%, 70%, 50%, 30%, 10%に減らしてCNNの学習を行い、認識精度への影響を検証した。結果を図6に示す。CNNでは大量のデータセットが高い認識精度を出すために必要とされているが、実験2ではゴッホの認識において枚数が減るごとに精度が下がっており、それが顕著に表れた。しかしレンブラント、モネの認識においては学習データが減ることで認識精度が上がる場合もある。これは普段の画風から外れた作品が減ったことで学習データがより最適化されたと考えられる。ただしこれはある特定の学習データセットを用いた場合に起こり、学習データを入れ替え

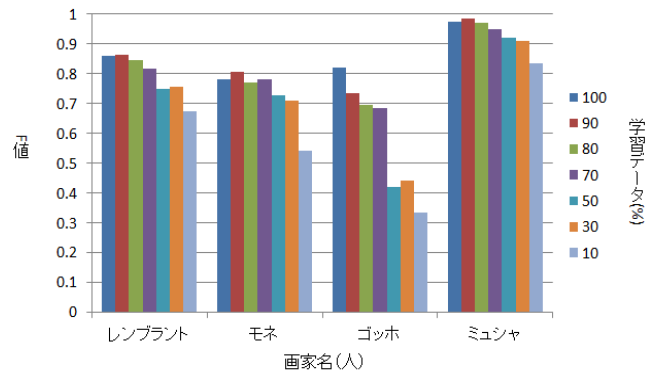


図 6: 学習データ減少による影響 (F 値)

ることで正規化を行った結果、その上昇値は低減し、全体として学習データの減少に伴って認識精度も低下した。画風にばらつきが多いほど学習データ量の減少により与えられる影響が大きい、ミュシャのような一定の画風を常に保ち続け、他とは大きく異なる画風の場合は、今回設定した学習データの減少でも精度低下の傾向は見られたものの他の画家ほど急激な変化ではなかった。

#### 4.5. 実験3 (画像分割方法の違いによる影響)

最後に、学習に用いる画像を分割する際に、オリジナルの画像の大きさにおいて32×32ピクセルのパッチ画像を切り出した場合と、リサイズを行ったうえで32×32ピクセルのパッチ画像を生成した場合の認識精度を比較した。図7はグループ1の認識をサイズ変更なしの場合とありの場合で行いそれぞれのF値を画家毎に比較したものである。サイズ変更を行うことによって全体的に僅かに精度が上昇しているのが読み取れる。画像分割の際に切り出しサイズの変更を行うことで、解像度のばらつきによる影響を抑えることができるが、同時に筆のタッチ情報が減少したものを学習してしまうという面もある。ミュシャの独特のタッチによる画風の認識精度が低くなったのはその影響と考えられる。上記のように精度が低下してしまう場合もあるが、局所的な特徴だけでなく全体の構図を意識した学習が行えたことで、ミュシャ以外の画家の分類精度は向上した。これら2つの例はトレードオフの関係に近いが、画家の人数が増えた場合、全体の構図を意識することはより重要になるためさらなる検証が必要だろう。

#### 5. 結論

本稿では、絵画画像を基にその絵画を描いた画家というラベルで分類した場合、どの画家に属するのかを推定し、さらに画風を特徴量として抽出した。画風の認識には近年一般物体認識をはじめ、音声認識や自然言語処理などの様々な分野で良い性能を示しているCNN

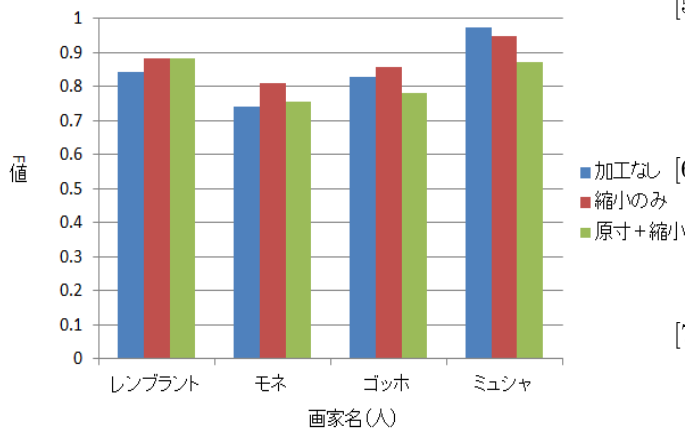


図7: 画像分割時のサイズ変更の有無による影響

を適用した。画家毎の分類では、他の画家と大きく異なった手法を用いている場合や全体として画風が統一されている場合は既存手法よりも良い精度で分類を行うことができた。また、学習データの減少によって認識精度が低下したことから、画像の分割による学習データ増加が精度の向上につながることを確認した。さらに画像分割の際に切り出しサイズに工夫を加えることで一部の画家の認識に効果があることを確認した。今回の実験では認識実験と精度の向上を目的としていたが、今後は学習した特徴を解析、応用することが課題である。ネットワーク構造の見直しや入力画像に他の工夫を施すことで認識精度のさらなる向上が見込める為、それらも検討したい。

#### 参考文献

- [1] Lior Shamir, Tomasz Macura, Nikita Orlov, D. Mark Eckley, Ilya G. Goldberg, "Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art", ACM Transactions on Applied Perception (TAP), Volume 7 Issue 2, Article No. 8, 2010.
- [2] Jonathan Jou, Sandeep Agrawal, "Artist Identification for Renaissance Paintings", [https://www.cs.duke.edu/~sandeep/research\\_files/artist\\_identification\\_fall2011.pdf](https://www.cs.duke.edu/~sandeep/research_files/artist_identification_fall2011.pdf), 2011.
- [3] Igor E. Berezhnoy, Eric O. Postma, H. Jaap van den Herik, "Automatic extraction of brushstroke orientation from paintings", Machine Vision and Applications, Vol.20, No.1, pp.1-9, 2009.
- [4] <http://www.image-net.org/challenges/LSVRC/2012/results.html>
- [5] Q. V. Le, M. A. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng, "Building High-level Features Using Large Scale Unsupervised Learning", ICML 2012.
- [6] Y. Bengio, A. C. Courville, and P. Vincent, "Unsupervised feature learning and deep learning: A review and new perspectives", CoRR, abs/1206.5538, 2012.
- [7] D. C. San-Cire, U. Meier, and J. Schmidhuber, "Multicolumn Deep Neural Networks for Image Classification", CVPR, 2012.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", NIPS, 2012.