

# 深層学習を用いた浮世絵に描かれた人物の人数自動カウント

傍士虎南<sup>†</sup> 生野壮一郎<sup>†</sup> 堀江和正<sup>‡</sup>

<sup>†</sup> 東京工科大学大学院バイオ・情報メディア研究科コンピュータサイエンス専攻

<sup>‡</sup> 筑波大学計算科学研究センター

## 1 はじめに

浮世絵は江戸時代に成立し、日本で起きた史実の内容を表しており、絵画の中で最も広く知られている絵画のジャンルの1つである。そのため、浮世絵内に描かれている内容を詳しく調べることは、史実の理解に直結するのである。しかしながら、浮世絵は画家によって技法、画風が異なり、また、合戦や祭事などの浮世絵内には描かれている情報量が多く、描かれている人物の人数を詳細に把握することは非常に時間と労力を費やしてしまう。

近年、深層学習は画像認識の分野で精力的に用いられている手法であるが、十分な認識精度を得るためには、学習データを十分に準備する必要がある。しかしながら、学習データとしての人物が数多く描かれた浮世絵は入手が困難である。

本研究の目的は、深層学習を用いて、浮世絵における人物に着目し、その人数を自動検出する手法を確立することである。また、十分な学習用のデータを確保するために、畳み込みニューラルネットワーク (Convolutional Neural Networks: CNN) を用いた画風変換技術 [1] と Deep Learning Convolutional Generative Adversarial Networks (DCGAN) [2] を用いて、学習データの水増しを行い、水増しデータが学習データとして機能するかを数値的に検証する。

## 2 学習データ

本研究では、深層学習を用いた物体検出分野の代表的な手法の1つである Faster R-CNN [3] を用いて学習を行っていく。この Faster R-CNN は、物体候補領域の検出にも深層学習を用いることで高精度化したニューラルネットワークである。

### Automatic count of persons in Ukiyo-e using Deep Learning

Konan Hoji<sup>†</sup> Soichiro Ikuno<sup>†</sup> Kazumasa Horie<sup>‡</sup>

<sup>†</sup>Tokyo University of Technology

192-0982, Tokyo, Japan

<sup>‡</sup>Center for Computational Sciences, University of Tsukuba

305-8577, Tokyo, Japan



図 1: 画風変換工程

上記のネットワークを用いて物体検出を行うために、必要な浮世絵の学習データを、立命館大学アート・リサーチセンターが所蔵するデータベース、浮世絵閲覧システム内の浮世絵画像データなどから人物が複数人描かれている絵を中心に収集を行う。これらの浮世絵をもとに、浮世絵内の人物を自動検出するための学習を行う。しかしながら、深層学習を行うには大量の学習データが必要となるが、学習に有用なデータは入手困難なことが問題としてあげられる。また、学習データとして足りない分のデータは、浮世絵閲覧システムから収集した浮世絵画像をもとに、画像生成技術である画風変換により画風を変換した画像と、DCGANにより作成した人物に似せた画像データの生成をそれぞれ行う。

### 2.1 画風変換

画風変換は図1のように、もとの1枚の画像(コンテンツ画像)に存在する物体の配置を維持しながらコンテンツ画像とは画風の違う別の画像(スタイル画像)の画風に変換する技術である。ネットワークの流れとしては、乱数で作成したノイズ画像を出力画像で用意し、ネットワーク内部にある損失計算用ネットワークにコンテンツ画像、スタイル画像、出力画像を入力し、出力画像とコンテンツ画像、スタイル画像を比べて損失を計算する。こうして計算された損失関数の和を損失関数を最小化する方向に出力画像を更新していくことで生成される。尚、各損失の計算、損失関数の詳細については参考文献 [1] を参照されたい。そのた

め、コンテンツ画像とスタイル画像の大きさは全て正方形で用意する必要がある。これは、長方形の画像を使用して画風変換を行うと、スタイルのズレが生じて画像内に空白ができてしまうことを避けるためである。

## 2.2 DCGAN

GANは生成器であるGeneratorと識別器であるDiscriminatorの2つのネットワークで成り立っており、このネットワーク同士を競合させることで学習データセットに用いた画像に近いデータを生成する。Generatorで学習データセット内の画像に酷似した画像を生成していき、Discriminatorで真偽の識別を行う。ここでは、真は学習データセット内の画像であり、偽は生成した画像とする。この真偽の識別を繰り返して行うことで本物に酷似した画像の生成を行うことができる。

本研究では、DCGANに用いる画像は、学習データとして用意した浮世絵画像から似た構図の人物を、正方形で出来る限り近い状態で複数切り出して、画像データの生成を行う。

## 3 実験

プログラムはPythonで実装し、ニューラルネットワークの構造はパッケージ: ChainercvのFaster R-CNNのサンプルを参考にネットワークを作成、学習を行う。実験に主に使用するパッケージの環境はchainer: 5.0.0 chainercv: 0.10.0を使用した。

### 3.1 実験方法

評価実験は学習中のmain/loss関数による正答率と学習を行ったモデルによるテスト実験によって評価を行う。また、本研究でFaster R-CNNの学習に使用される学習データは、本来の浮世絵画像、画風変換を用いて生成した浮世絵スタイルの画像データ、DCGANを用いて生成した画像データの3種類で構成されている。これは、本来の浮世絵画像のみでは学習が十分に行えないと判断したためである。また、2種類の画像生成技術を用いたのは、先に触れたDCGANでは有用なデータが生成しづらかったためである。

### 3.2 実験結果

学習中の損失関数: lossの動きを図2に示す。図2に示すように、学習中のlossがほぼ0となる結果であった。しかしながら、生成したモデルを用いてテストをした結果、学習に用いた画像データの正答率はほぼ100%となったが、新しい画像を用いてテストを行った場合には半分も検出することが出来なかった。

テストの結果が乏しい要因として、過学習の影響と

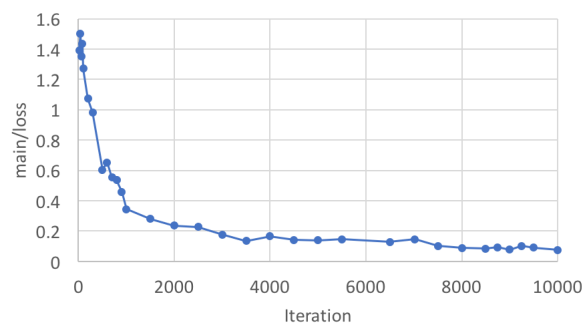


図 2: 学習結果

学習データ自体が不足していることが考えられる。そのため、学習回数を見直しつつ、学習データを増やすことが出来れば実用化は可能であると考えられる。

## 4 おわりに

本研究では、深層学習を用いて、浮世絵における人物の人数を自動検出する手法を確立すると共に、畳み込みニューラルネットワークを用いた画風変換技術とDCGANを用いて学習データの水増しを行い、水増しデータが学習データとして機能するかの検証を行った。

実験の結果、学習データにあるデータであれば十分に検出は可能であることが確認できた。本実験により、鎧を着た人物や傘を被っている人物などの顔が見えない状態である人物であっても、学習データの方で十分に補うことは可能であると考えられる。

## 参考文献

- [1] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge, "A Neural Algorithm of Artistic Style" in arXiv:1508.06576v2 [cs.CV], 2015.
- [2] Alec Radford, Luke Metz, Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" in arXiv:1511.06434 [cs.LG], 2016.
- [3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards Real-Time object detection with Region Proposal Networks" in arXiv:1506.01497v3 [cs.CV], 2016.