

図書館における機械学習技術の実践的適用について

次世代デジタルライブラリーの機能改修及び新たなデータセットの公開を中心に

青池 亨 (国立国会図書館)

概要: 国立国会図書館のデジタル化資料について、機械学習を適用したサービス提供の改善を試みるに当たっては、目指す機能を実現するための有効性や構築コストに考慮してデータセットの設計・整備を行うこと、また、既存の大規模データ資源や目覚ましく進展する画像認識分野の機械学習アルゴリズムを活用することの両方に目を配る必要がある。本論文では、新たに構築・公開したデータセット2件（画像タグデータセット NDL-ImageLabel 及び OCR1 行データセット）について、その設計思想と有効性を論じ、実際の活用事例として2021年末にサービスの改修を予定している次世代デジタルライブラリーの技術要素や新規事業の技術検討における役割を紹介する。本論文で取り上げたデータセットや機械学習モデルについてはNDLラボのGitHub (<https://github.com/ndl-lab/>) から公開している。

キーワード: データセット, 機械学習, 画像検索, OCR, 次世代デジタルライブラリー

On the practical application of machine learning in the library

Focusing on the functional modification of the next digital library and the release of new datasets

Toru Aoike (National Diet Library)

Abstract: When attempting to improve services by applying machine learning to the digitized materials held by the National Diet Library (NDL), it is important to consider the preparation of data sets based on a design that takes into account the effectiveness and construction cost to realize aimed functions. It is also necessary to pay attention to the coordination of existing large-scale data resources and machine learning algorithms in the field of image recognition, which has made remarkable progress in recent years. In this paper, we discuss the design concept and effectiveness of two newly released datasets (image tagging dataset NDL-ImageLabel and OCR one-line dataset) and, as an actual case study, introduce how they are playing a role both in the in-house technical renovation of the Next Digital Library, which is scheduled to be released to the public at the end of the FY2021, and in the technical investigation for new library services. The datasets and the machine learning models discussed in this paper are available on GitHub repository of NDL Labs (<https://github.com/ndl-lab/>).

Keywords: Dataset, Machine Learning, Image Retrieval, OCR, Next Digital Library

1. 背景

国立国会図書館電子情報部電子情報企画課次世代システム開発研究室（次世代室）では、機械学習技術を図書館サービスに取り入れ、応用することを目的に調査研究活動を行っており、研究成果を活用したサービスを一般に利用可能な形で実験的に提供する場として、2019年3月に「次世代デジタルライブラリー (<https://lab.ndl.go.jp/dl/>) [1]」を公開した。

一般公開から2年が経過した次世代デジタルライブラリーであるが、職員の手で構築・公開・運用することによって得られた知見（機械学習技術、OCR テキストデータに対する検索技術、サービスを実現するためのシステムインフラ構築における技術的知見）は、当館が実施する他の事業に活用されてきた。事例を挙げて紹介すると、当館

が運用する既存サービスの新規機能開発（例えば、ジャパンサーチの画像検索機能）、職員による作業の効率化ツールへの転用（例えば、資料デジタル化委託作業における納品データの検品時における撮影位置の自動チェックツール）、研究開発を外委託する際の仕様検討や受託者への情報提供（例えば、令和3年度に実施中のOCR処理プログラム研究開発作業における検討）の基礎となった。他方で、次世代デジタルライブラリーの機能は、検討当時の機械学習技術を採用したものであり、技術的陈腐化が進んでいる。当館における先端的技術導入のためのパイロット事業としての価値を維持するためには、手法の刷新と新しい実験に取り組むためのデータセットが必要である。

さらに、当館以外の図書館や研究機関においても、画像検索やレイアウト認識等、類似したサー

ビスの研究やデータセットの提供が進んできた。例えば、米国議会図書館が開発し、2020年に公開した Newspaper Navigator[2]は、1900年から1963年までに発行された新聞の資料画像に対して画像で検索したり、OCR テキストに対してキーワードで検索したりすることが可能なサービスである。このサービスは、データセットやソースコードも公開されている。また、The iMet Collection 2019 Challenge Dataset (iMet)[3]は、メトロポリタン美術館の所蔵する美術品について、その材質や形状、製作された文化圏をタグ付けしたデータセットである。このデータセットにより、美術品画像への機械学習の適用が容易になった。

国内のデジタルアーカイブにおける自動タグ付けの先行事例としては、国立情報学研究所が「華北交通アーカイブ」[4]において2019年に公開したタグ付与機能がある。華北交通アーカイブでは、一般物体認識のデータセットである ImageNet を学習したモデルによる、1000種類のタグを画像毎に自動で付与している。また、人文学オープンデータ共同利用センターの提供する顔貌コレクション(<http://codh.rois.ac.jp/face/>)の顔コレデータセット[5]は、日本の絵巻物や絵本に登場する様々な顔貌について領域を切り出し、身分や性別といった描かれた顔貌の属性について機械学習による自動タグ付けの学習データとして利用できるようなメタデータを整備したデータセットである。

これらの研究を分析し、有効性が高いと考えられる手法やデータセットの活用に向けて、未着手の領域へのアプローチを含めた検討を行い、これらの検討結果をデータセットやオープンソースの形態で公開していくことは、他の図書館等における最新技術の導入を手助けすることにとどまらず、我が国のオープンデータ利用促進の観点からも意義があると考えられる。

2. データセットの作成と公開

当館のデジタル化資料を利用した画像認識及びテキスト化に当たっての調査と実験準備を兼ねて、当館が保有するデジタル化資料画像を加工して、画像タグデータセットと OCR1 行データセットの2種類のデータセットを作成した。

2.1. 画像タグデータセット (NDL-ImageLabel)

NDL-ImageLabel は、次世代デジタルライブラリーの API を利用し、著作権保護期間の満了した資料から自動抽出された挿絵や写真(以下、まとめて「図版」という)の画像に対して、人手で分類を行って作成したデータセットである。内訳は表1の通りである。

表 1 NDL-ImageLabel の構成

| タグ名称 | 画像数 | 内容 |
|----------------------|-------|----------|
| graphic_map | 581 | 地図 |
| graphic_graph | 371 | グラフ資料 |
| graphic_illustrator | 170 | カラーイラスト |
| graphic_illustration | 2,695 | モノクロイラスト |
| picture_landmark | 728 | 屋外建造物写真 |
| picture_outdoor | 748 | 屋外写真 |
| picture_object | 401 | 美術品等写真 |
| picture_indoor | 232 | 屋内写真 |
| picture_person | 499 | 肖像写真 |
| stamp | 479 | 印影・蔵書票 |

picture_person 以外の各タグに該当する図版の例を図1から図9までに示した。

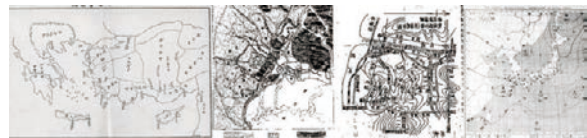


図 1. graphic_map(地図)の例

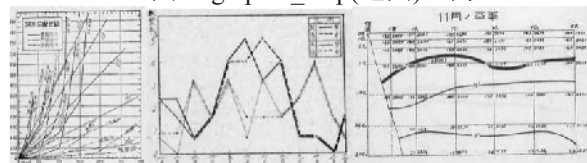


図 2. graphic_graph(グラフ)の例



図 3. graphic_illustrator(カラーイラスト)の例



図 4. graphic_illustration(イラスト)の例



図 5. picture_landmark(屋外建造物写真)の例



図 6. picture_outdoor(屋外写真)の例



図 7. picture_object(美術品等写真)の例



図 8. picture_indoor(屋内写真)の例

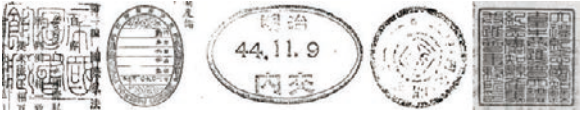


図 9. stamp(印影・蔵書票)の例

肖像権への配慮から除外した picture_person 以外の 9 種類については、次の URL から 2021 年 7 月に公開している。

<https://github.com/ndl-lab/imagetagdataset>

本データセットは、地図やグラフ資料、印影等、資料内の掲載箇所を特定できると図書館資料の利用の効率化に資することが期待される分類タグを設定しているほか、屋外の地点の目印となる建造物等を指すランドマーク (Google Landmark Dataset v2[6]) や美術品 (iMet) や屋内写真 (Indoor Scene Recognition[7]) といった各分野において先行するオープンな画像データセットに関連付けを行える構成とした。

今回の構成に至った理由は、データセット構築時に次に挙げる課題があることを認識したためである。

資料に含まれる図版は多様であり、ImageNet[8] のようにこれらを網羅して説明する詳細なタグ集合を新たに設計するアプローチをとることは現実的ではない。また、解像度や色数について一般的な機械学習に用いる画像より大きく劣るため、仮に件名標目や WordNet 等既存の詳細な階層設計に基づいてタグ付けを行った場合、付与すべき正確なタグを判定できない画像や学習に必要な数量を確保できないクラスが多く発生し、大きなコストがかかることが想定される。

もう一つのアプローチとして、先行研究の華北交通アーカイブの事例のように一般物体認識のタスクに最も汎用的に利用されている ImageNet によるクラス分類や一般物体認識用の API の出力するラベルをそのまま用いる方法が考えられる。しかし、ImageNet の 1000 クラスのうち 4 割程度について動物を分類するクラスが占めており、今回対象としている当館デジタル化資料の資料画像がもつ多様性に対して必ずしも適合するものではない。また、Google Cloud Vision のような画像から物体を認識してタグ付けする API を事前のトレーニングなしで利用した場合、例えば

graphic_map(地図)に該当する地図資料の多くは「animal(動物)」のタグ、graphic_graph(グラフ資料)に該当する資料では「license_plate(ナンバープレート)」のタグが付与される等、認識対象としてカバーされていない種類の画像について、学習済みのタグに無理やり当てはめようとするために誤った推定結果が付与されてしまう(図 10)。

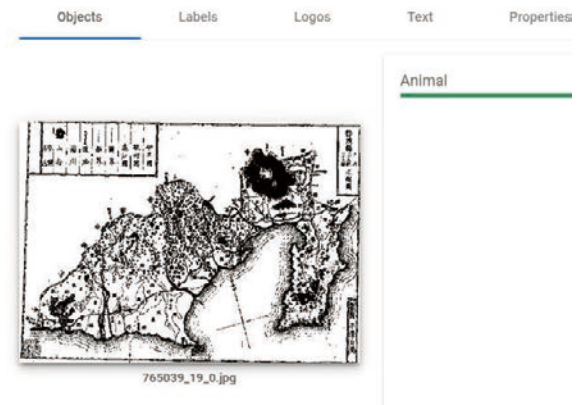


図 10. Google Cloud Vision API により、古地図(現在の静岡県)に Animal のタグが付与される例

こうした状況を避けるためには、資料画像から切り出される図版を、第一段階としては大きな粒度でも良いので取りこぼしの少ない、網羅性の高い単位で切り分けて分類し、第二段階として切り分けられたそれぞれの分野に特化した画像データセットによる機械学習を行うことが望ましいと考えられる。

上記を踏まえ、NDL-ImageLabel は、これらのアプローチの短所を補うことを目的に、多様な資料画像の特性を振り分けることを第一段階の機械学習処理とすると、既に研究の進んでいる分野において存在する大規模な画像データセットを活用した機械学習を適用した第二段階の処理にスムーズにつながるよう、橋渡しの役割を果たすデータセットとなることを目指した(図 11)。すなわち、既存のデータセットがカバーしている分類については、既に先行研究の存在する詳細かつ大規模なデータセットを利用した機械学習モデルに引き渡しやすいように設計している。

これにより、機械学習のために先行研究として整備されてきたオープンかつ大規模なデータ資源と当館のパブリックドメインな資料画像のデータ資源の間をつなぎ、より広範な用途への活用につなげることができると考える。

3.1 及び 3.2 にて実際に本データセットと大規模データ資源を連携して活用して行った次世代デジタルライブラリーの機能改修について紹介する。

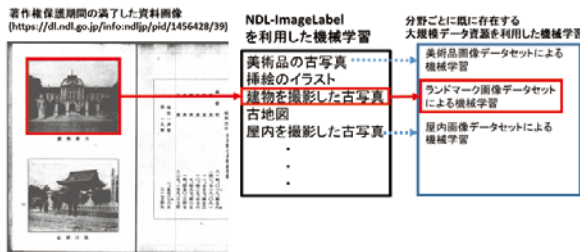


図 11. NDL-ImageLabel が当館デジタル化資料の資料画像と各分野に存在する大規模データ資源との間を接続するイメージ(例えば、資料から切り出された図版が建物を撮影した写真であることを自動判定できれば、このような図版のみに対して選択的に、ランドマーク画像検索用の大規模データセットを学習に利用した高性能・高機能な機械学習モデルと連携できる。)

2.2. OCR1 行データセット

当館デジタル化資料を提供する正式サービスである「国立国会図書館デジタルコレクション」では、デジタル化資料の画像や書誌データのほか、利便性のため資料の目次情報のテキストデータを提供している。この目次情報のテキストデータは、外部委託により人手で作成したものであるため、目次が含まれる資料画像については、概ね高精度なテキストデータを保有していると言える。一方で、資料画像中のどの座標に対応するテキストデータが存在するか、という情報は持っていない。このため、両者を関連付ける作業を行うことによる OCR 学習用データセットの構築について検討した。

具体的な作業としては、画像中の目次の文字列領域を行ごとに切り出し、目次情報のテキストデータから合致するテキストの部分をはめることで、1 行ごとに切り出された文字列画像と、文字列画像中のテキストデータの対応セットを作成した。

本データセットについては著作権保護期間の満了した資料から作成したサブセットを次の URL から公開している。

<https://github.com/ndl-lab/ocr-ndloneline>

3. データセットを利用した調査研究

3.1. NDL-ImageLabel を利用した自動タグ付与手法の再検討

現行の次世代デジタルライブラリーにおいても、タグの自動付与機能を試験的に提供している。しかし、学習対象となるデータセットの設計構築に時間が確保できなかったことから、あくまで試行的な付与にとどまり、精度やタグセットの設計

に課題があった。今回作成した NDL-ImageLabel を用いて学習を行い、タグ付与と精度の改善を図った。

モデルには EfficientNetB0[9]を用いた。

モデル及び学習・推論のコードについては次の URL から公開している。

<https://github.com/ndl-lab/tagestimatemodel>

3.2. 画像検索向け特徴抽出手法の再検討

次世代デジタルライブラリーの画像の特徴抽出には、これまで ImageNet で事前学習された MobileNet V2[10]の Global Average Pooling 層の出力を利用していたが、出力次元数が 1280 次元と大きく、インデックスの保持に多くのメモリを要するため、サービスを提供する上で、高性能なサーバの確保が必要という課題があった。また、ImageNet は画像の殆どがカラー写真からなるデータセットであり、次世代デジタルライブラリーが主な検索対象とするグレースケールの画像に対して特化した学習を行っていなかった。上記の点を踏まえて、NDL-ImageLabel によるモデルと組み合わせることを前提に、資料画像中の汎用的な写真検索のモデル、美術品検索用途のモデル、屋外画像検索用途のモデルの 3 種類を作成した。モデル及び学習・特徴ベクトル出力のコードは、次の URL から公開している。

<https://github.com/ndl-lab/imageretrivalmodels>

3.2.1 写真検索及び美術品検索用途のモデル作成と可視化

EfficientNet B0 をバックボーンに出力次元数を 512 次元とした画像認識モデルを作成し、グレースケール化したデータセットに対して距離学習を利用した転移学習を行った。

転移学習対象となるデータセットは、Google 社の提供するオープンな一般物体認識画像データセットである Open Images Dataset V6[11](OID)及びメトロポリタン美術館の提供するオープンな美術品画像データセットである iMet を利用した。(表 2)

表 2 学習に利用したデータセットの情報

| データセット名 | クラス数 | 画像数(train のみ) |
|---------|-------|---------------|
| OID | 601 | 1,743,042 |
| iMet | 1,103 | 109,230 |

これらはひとつの画像に対して複数のラベルが付与されることの多いマルチラベルなデータセットであるため、学習対象の損失関数には binary cross entropy with logit を利用している。

距離学習は、学習時に同一のラベル同士はより近く、異なるラベル同士はより遠くなるよう制約

を設けることで、類似検索に適した特徴を得るための手法である。本研究では EfficientNet B0 に結合した 512 次元の全結合層に対して L2 Softmax[12]を適用した。

実験は、表 3 に記載した 4 パターンで行った。比較のため、②から④までは、データセット間で異なるクラス数を除く学習時のパラメータについて全て揃えている。

表 3 利用したモデルと学習対象の組み合わせ

| No | モデル名称 | 事前学習 | 転移学習 |
|----|----------------|----------|----------|
| ① | MobileNetV2 | ImageNet | - (従来手法) |
| ② | EfficientNetB0 | ImageNet | OID |
| ③ | EfficientNetB0 | ImageNet | iMet |
| ④ | EfficientNetB0 | OID | iMet |

評価方法：

NDL-ImageLabel の graphic_illustrator を除く全データを対象として、各画像から得られた特徴ベクトル同士の位置関係を実験条件ごとに調べた。本論文では可視化のため、UMAP[13] (近傍数 5, Spectral Clustering) を用いて、二次元座標上にマッピングした。

結果：

①から④までの可視化結果を図 12 から図 15 までにまとめた。検索結果の最適化の観点では、同一タグが付与された画像同士が一定の集合を形成し、似た関係にあるタグ (例えば紫色の picture_landmark とピンク色の picture_outdoor) について、集合同士が近接していること、また他のタグと性質の異なるタグ (例えば灰色の picture_person) の集合が他の集合と離れて配置されることが望ましい。①では同一タグが付与された画像群が離れた複数の集合に分かれるものや、他の画像と孤立して配置された画像も見られ、ドメインに対するチューニングの必要性を示唆するものとなった。②では同一タグごとにまとまった集合を形成するものの、タグ間の差異がやや曖昧となっている。③ではタグ間の差異が強調された。④では②と③の中間的な配置となった。以上の可視化結果から、特に美術品に関しては④のモデルを用いて抽出した特徴が、全般的には③のモデルを用いて抽出した特徴が適すると考えられた。

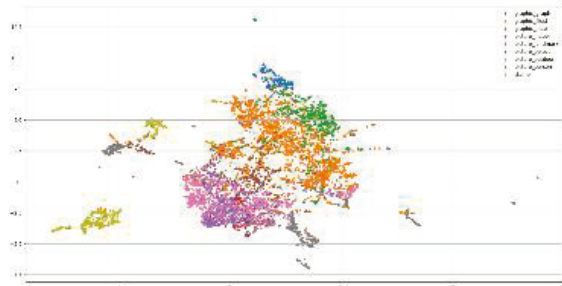


図 12 ImageNet で学習した MobileNetV2 による特徴ベクトルの可視化 (表 3①)



図 13 OID で学習した EfficientNetB0 の特徴ベクトルの可視化 (表 3②)

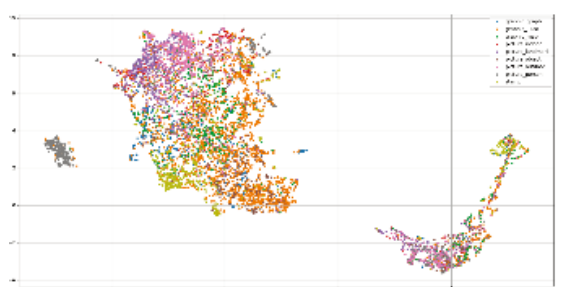


図 14 iMet で学習した EfficientNetB0 の特徴ベクトルの可視化 (表 3③)

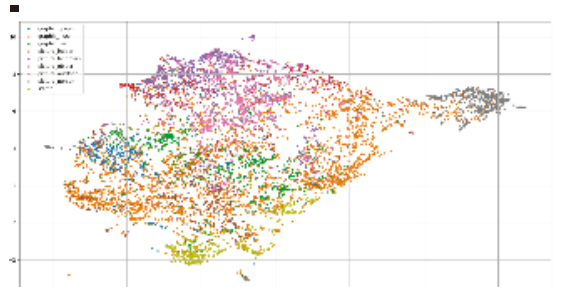


図 15 OID で事前学習し、iMet で転移学習した EfficientNetB0 の特徴ベクトルの可視化 (表 3④)

3.2.2 大規模データ資源を利用した屋外ランドマーク検索用途の高精度画像検索モデルの作成

3.1 で述べた自動タグ付与モデルを介して適用対象を絞り込むことによって、既存の大規模なデータ資源や有力な研究成果が存在する領域に対して有効な高性能な特化型モデルを構築し、資料検索性能を向上させるアプローチを試みた。

Google Landmark Dataset v2[6](GLDv2)を学習の対象として、屋外の建造物等の画像検索に特化したモデルを構築した。

学習に利用したデータセットは、GLDv2に含まれる約20万地点、4,132,914画像から検索対象となるランドマークの数を81,313地点に限定し、ノイズとなる画像を取り除いた1,580,470画像からなるデータセット(GLDv2と同時に提供され、cleanバージョンと称される)をベースとし、これに20画像未満のクラスが20画像になるように不足分をGLDv2から補った合計1,985,432画像からなる。

モデルにはEfficientNetv2-S[14]を利用し、距離学習には人間の顔認証を行うための学習に利用されるArcFace[15]を用いている。学習の過程でモデルの性能が収束するごとに、入力画像サイズを徐々に、224x224→360x360→480x480→600x600の4段階で大きくし、また、Data Augmentationを強めながら(random resize 及び random cropを画像幅に対して5%→10%→10%→15%に強める)学習を行った後、抽出図版の解像度に合わせて入力画像サイズを224x224に落として追加の学習を行った。特徴ベクトルの次元数は3.2.1のモデルと同様に512次元である。

なお、本モデルに追加修正を加えた機械学習モデルにより、著者はGoogle Landmark Retrieval2021[16](GLDv2を利用して画像検索用の機械学習モデルを作成し、ランドマーク画像に対する検索性能を競うGoogleが毎年1回開催している国際コンペティション。評価用データには[17]のデータセットが用いられる)の大会において第19位を獲得している。相違点は、「入力画像サイズを落とさず、600x600のままとしている点」「foldの異なる4モデルの出力ベクトルを連結して2048次元の特徴ベクトルとしている点」「後処理としての再ランキング処理を追加している点」の3点であり、本論文の実験ではいずれも大量の資料画像を処理する上で軽量化を図るため、採用しなかった。

3.3. レイアウト認識処理手法の再検討

次世代デジタルライブラリーでは、資料画像中の図版の領域を自動抽出し、類似した図版を探す画像検索機能を提供している。これまではsemantic segmentationの手法を元に、図版と推定された領域の外接矩形を取ることでこの機能を実現してきた。2019年3月公開時には、日本十進分類法(NDC)5類から7類までを対象としており約6万資料(600万画像程度)と少なかったため、比較的計算コストの大きな手法であるDeepLabv3 plus[18]モデルを利用した。その後、

2020年8月に著作権保護期間満了資料の全件を対象とすることで対象資料数が約5倍の33万資料(3,300万画像程度)と増加したため、モデルをより軽量のESPNetv2[19]に変更し、1画像あたり300ms程度まで高速化を行った。一方、令和5年度以降にリリース予定の次期の国立国会図書館デジタルコレクションでは、当館が保有する全てのデジタル化資料に対して、同様の自動抽出機能を提供する予定であるため、240万資料(2.3億画像程度)に対して事前に検出処理を行う必要がある。このため、抜本的なアルゴリズムの見直しを図った。semantic segmentationで物体検出を行う欠点として、本来ほぼ直線の組み合わせで表現可能なレイアウトに対して不必要に自由度の大きな推定を行うため、計算に無駄がある点、撮影時の光源により生じた陰影のために推定された境界の一部が不明瞭になり、外接矩形が本来の位置からずれる点が挙げられる。

これらの欠点を解決し、更なる高速化を図るため、公開しているNDL DocL[20]を学習させたYOLOv5[21]によって自動認識を行うアプローチに切り替えた。従来手法との認識性能の比較を表4に示した。

表4 学習に利用したモデルと認識性能の比較

| モデル名称 | 資料領域 認識性能 (mean IOU) | 図版検出 性能 (mAP@0.3) | 処理速度 (1画像あた りの概算) |
|------------|-------------------------------|-------------------------|-------------------------|
| DeepLabv3+ | 0.65 | 0.37 | 800ms |
| ESPNetv2 | 0.89 | 0.66 | 300ms |
| YOLOv5 | 0.96 | 0.62 | 100ms |

注：処理速度について、以下のハードウェアによる実行環境を想定

- Intel(R) Xeon(R) W-3245 CPU @ 3.20GHz 1基
- GPU:NVIDIA Geforce RTX 2080Ti 1基

これにより、1画像当たり1処理当たりの計算時間を100ms程度まで圧縮することができたため、並列処理によって目的を達成できる見通しとなった。著作権保護期間満了資料全てに対して、実際に処理を行い、概ね概算通りの期間で処理が確認することを確認した。この確認作業の過程で、副産物として、次世代デジタルライブラリーが検索対象とする資料全件の図版抽出結果が得られたため、次世代デジタルライブラリーの挿絵抽出処理の機能を差し替えた。

4. まとめ

本論文では、まず当館のデジタル化資料の資料

画像を利用した機械学習に関する調査研究への理解をより深めてもらうための準備として、資料画像から抽出した図版のタグデータセットである NDL-ImageLabel と目次を使った OCR1 行データセットの構築について、その設計方針と意義について説明した。次に、実際に次世代デジタルライブラリーのタグ付与精度や画像検索精度を改善するためのモデル作成を行った過程を紹介した。モデル作成の過程で、NDL-ImageLabel を活用することで、画像検索用の特徴量がどのような性質を持つか可視化する手法を示した。また、NDL-ImageLabel を利用して多様な図版の中から改善したいタグを絞り込むことによって、既に大規模なデータセットの構築が完了しており、性能を競うコンペティションが定期的開催されるなど研究が盛んなランドマーク画像検索の分野で開発した高性能なモデルが適用できるようになり、次世代デジタルライブラリー上で、屋外建造物の図版の画像検索精度を大きく改善できることを示した。

また、資料画像中から図版を切り出す機械学習手法についてもより大量の資料画像から精度を維持して高速に抽出できるよう、手法の見直しを行った。これにより大量の資料画像から容易に図版を取得できるようになり、画像タグデータセットのより一層の改良や今後のデータ量拡張を容易に行えるようになった。

5. 今後の展開について

令和3年度、次世代室は当館所蔵の図書・雑誌等のデジタル化資料の全件 OCR テキスト化と、オープンなライセンスで提供可能な OCR 処理プログラムの研究開発の2つの事業を外部委託にて実施している。これらは、第一義的には国立国会図書館デジタルコレクションにおける全文検索機能の実装や視覚障害者等用読み上げテキストの提供を念頭に置いた事業ではあるが、副次的に膨大な量の未校正 OCR テキストデータと、当館が設定した性能水準を満たすために作成された OCR 学習用データセットを入手できる。入手でき次第、次世代デジタルライブラリーの全文検索の対象範囲を、NDC6 類のみから全ての著作権保護期間満了資料に拡大する予定である。今後の国立国会図書館デジタルコレクションでのサービス拡充の際の参考材料を増やすため、より広くフィードバックを求めている。

先端技術を活用したよりよい図書館サービスの検討のため、これまでの画像を中心とした調査研究及び実験事業に加えて、固有表現抽出等の自然言語処理に関する調査研究活動も行っていく

必要があると考えている。共同研究等、ご関心のある研究者の皆様の協力を得ながら進めたい。

6. 謝辞

OCR1 行テキストデータの作成作業に当たっては、当館職員の辻真理子氏に多くの協力を頂いた。また、いつも有益な議論と助言を頂いている次世代室の皆さんにこの場を借りて感謝の意を表す。

参考文献

- [1]青池亨, 里見航, 川島隆徳. "資料画像中の挿絵領域の自動抽出及び画像検索システムの実装." *じんもんこん 2018 論文集* (2018): 97-102.
- [2]<https://news-navigator.labs.loc.gov/search>
- [3]Zhang, Chenyang, et al. "The iMet collection 2019 challenge dataset." *arXiv preprint arXiv:1906.00901* (2019).
- [4]<http://codh.rois.ac.jp/north-china-railway/>
- [5]Tian, Yingtao, et al. "Kaokore: A pre-modern j apanese art facial expression dataset." *arXiv preprint arXiv:2002.08595* (2020).
- [6]Weyand, Tobias, et al. "Google landmarks data set v2-a large-scale benchmark for instance-level r ecognition and retrieval." *Proceedings of the IEEE /CVF conference on computer vision and pattern recognition*. (2020).
- [7]Quattoni, Ariadna, and Antonio Torralba. "Reco gnizing indoor scenes." *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, (2009).
- [8]Deng, Jia, et al. "Imagenet: A large-scale hiera rchical image database." *2009 IEEE Conference o n computer vision and pattern recognition. Ieee*, (2009).
- [9]Tan, Mingxing, and Quoc Le. "Efficientnet: Re thinking model scaling for convolutional neural n etworks." *International Conference on Machine Le arning*. PMLR, (2019).
- [10]Sandler, Mark, et al. "Mobilenetv2: Inverted r esiduals and linear bottlenecks." *Proceedings of th e IEEE conference on computer vision and patter n recognition*. (2018).
- [11]Papadopoulos, Dim P., et al. "Extreme clickin g for efficient object annotation." *Proceedings of the IEEE international conference on computer vi sion*. (2017).
- [12]Ranjan, Rajeev, Carlos D. Castillo, and Rama Chellappa. "L2-constrained softmax loss for discri minative face verification." *arXiv preprint arXiv:1703.09507* (2017).
- [13]Lopes, António M., and José A. Tenreiro Ma chado. "Uniform Manifold Approximation and Pro jection Analysis of Soccer Players." *Entropy* 23.7 (2021): 793.

[14]Tan, Mingxing, and Quoc V. Le. "Efficientnet v2: Smaller models and faster training." *arXiv preprint arXiv:2104.00298* (2021).

[15]Deng, Jiankang, et al. "Arcface: Additive angular margin loss for deep face recognition." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2019).

[16]<https://www.kaggle.com/c/landmark-retrieval-2021>

[17] Kim, Zu, et al. "Towards a fairer landmark recognition dataset." *arXiv preprint arXiv:2108.08874* (2021).

[18]Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." *Proceedings of the European conference on computer vision (ECCV)*. (2018).

[19] Mehta, Sachin, et al. "Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2019).

[20] 青池亨, 木下貴文, 里見航, 川島隆徳. "機械学習のための資料レイアウトデータセットの構築と公開." *じんもんこん 2019 論文集* (2019): 115-120.

[21] <https://github.com/ultralytics/yolov5>