N-041

画像内の展示物と説明パネルの関係性推定に関する研究
A study on Estimating Relationships between Multiple Exhibits and Description Panels in a Single Image

志田 晃一† 赤嶺 有平‡ 根路銘もえ子§
Koichi Shida Yuhei Akamine Moeko Nerome

1. 研究背景と目的

近年の通信環境の整備やモバイル端末の普及に伴うデジタル化の加速により,インターネットを介した情報のやり取りや公開の機会が増加している.また,新型コロナウイルス感染症蔓延時に博物館施設の利用が制限された経験から,博物館においても資料をデジタルアーカイブ化してインターネットを通じて公開することは,資料の保存と活用の観点からもその重要性が認識されている.

2023 年に施行された博物館法の改正において,博物館資料のデジタルアーカイブ作成およびその公開が博物館の事業として位置づけられたことにより,デジタルアーカイブ化の更なる加速が予測される.一方で,博物館の展示物とそれに関連する情報をデジタルアーカイブとして保存するには,博物館に関する専門知識だけでなくデジタルアーカイブに関する専門知識が必要であるが,必要な人材を大小全ての博物館が独自に確保・育成することは困難である.令和 2 年度に実施された調査によると,デジタル技術を活用した取り組みを実施する上で,予算の不足・人員の不足・知識ノウハウの不足が 56~59%の博物館で課題として挙げられ,73.6%の博物館でデジタルアーカイブに関する専門知識を持った職員が在籍していない状況であることが示されている[1].この調査結果から,多くの博物館が専門性を持たないままデジタルアーカイブ化に取り組んでいる状況であることがわかる.また,人材だけでなく予算の不足も課題としてあげられている.従来のデジタルアーカイブ化手法では展示物を 1 つずつ撮影し,展示物に関連する情報を紐付けて展示物のデジタルアーカイブ作成を行なっている.しかし,展示物を個別に撮影してデジタルアーカイブを作成する場合,膨大な時間と人員が必要であり,そのような手法を大小全ての博物館で実施することは困難である.

そこで,既に展示されている展示物とそれに関連する情報のデジタルアーカイブ化を,専門知識を必要とせずに安価に実現する手法の確立に取り組んでいる[2].本研究で取り組んでいるデジタルアーカイブ化手法の流れを図 1 に示す.既に展示されている展示物とパネルの情報を自動で取得するために,自動走行ロボットを用いた展示物の撮影を行う.自動走行ロボットが取得した画像から,展示物とパネルの検出を物体検出モデルを用いて行う.検出されたパネルに記載されている説明文に対して OCR による文字認識を行うことで,展示物の画像と解説をまとめたデータセットを構築する.

† 琉球大学理工学研究科 Graduate school of Engineering and science, Ryukyu University
‡ 琉球大学工学部 Faculty of Engineering , Ryukyu University
§ 沖縄国際大学経済学部 College of Economics and Environmental Policy, Okinawa International University

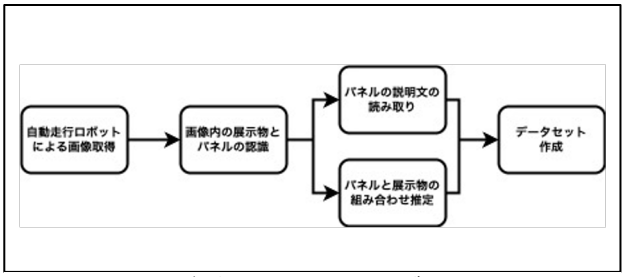

図 1 デジタルアーカイブ化の流れ

自動走行ロボットで撮影される画像は,展示されている展示物を 1 つずつ個別に撮影したものではなくテーマごとにまとめて展示されている展示物を撮影することを想定しているため,画像内には複数の展示物とそれに関連する説明パネルが写っている.そのため,検出された複数の展示物と説明パネルの組み合わせを正しく行なった上でデータセットを作成しなければ,展示物に対して誤った情報が紐づけられたデータセットを作成してしまう.

本研究では,検出された複数の展示物と説明パネル間の適切な組み合わせを推定する手法についてルールベース手法とニューラルネットワークベースのモデルによる手法を提案し,デジタルアーカイブ自動化システムの構築に貢献することを目指す.

2. 関連研究

本研究が対象とする画像内の展示物と説明パネル間の関係性推定は,画像内の物体間における意味的関係を抽出するタスクとして知られるシーングラフ生成（Scene Graph Generation: SGG）と関連する.SGG は画像内の物体（エンティティ）をノード,物体間の関係をエッジとして表現するグラフ構造を生成するものである.

SGG 手法の代表例として,Xu ら[3]は各構成要素となる物体と関係性を独立に推論するのではなく,文脈情報を活用した共同推論によって反復的に予測を改善する手法を提案した.このアプローチでは,物体候補領域から抽出した視覚的特徴をグラフ推論モジュールに入力し,反復的にノードとエッジを更新する.そして,最適化後のノードとエッジの状態に基づいて物体カテゴリ・ペア間の関係性タイプを予測する.これにより,複数のベンチマークデータセットで従来手法を上回る精度を達成した[3].

既存の SGG モデルは,豊富な関係性クラス分類を学習しており,モデル評価に用いた Visual Genome[4]データセットでは関係性クラスが 50 種類存在する.一方で,本研究で取り扱うタスクは,展示物と説明パネル間の関連の有無を予測する二値分類問題であり,既存の SGG モデルは過剰な構造を持つと言える.

以上のことから,本研究では SGG の基本的な考え方を踏まえつつも,博物館の展示物と説明パネル間の関係の有無推定という,ドメインに特化した二値分類タスクに適した手法を提案する.

3. 提案手法

本研究では,博物館における展示物と説明パネルは,閲覧者にとってその関連性が直感的に理解しやすいよう,視覚的・空間的に工夫された配置がなされているという前提から,展示物と説明パネル間の関係の有無は,両者の位置関係や大きさといった幾何的情報に強く依存しているという仮説に基づき,ルールベース手法,1 組の展示物と説明パネル間の関係を推定するニューラルネットワークモデル,1 つの展示物と画像内の複数の説明パネル間の関係を推定するニューラルネットワークモデルの 3 つの手法を提案する.

本研究で対象としているタスクの適用例を図 2 に示す.図 2 の左の画像は,博物館の展示物を撮影した画像に対して物体検出により展示物と説明パネルを検出した画像である.展示物は黄色い矩形で囲まれ,説明パネルは緑色の矩形で囲まれている.図 2 の右の画像は,本研究が取り組む展示物と説明パネルの関係性推定を行った結果を示した画像である.左の画像に青色の直線が追加されており,これは展示物と関係がある説明パネルとの間に直線を描いている.この画像では,6 個の展示物はすべて中心にある説明パネルと関係性があると予測されることを目指す.

3.1 ルールベース手法

ルールベース手法では,以下の条件を満たすパネルを展示物と組み合わせる.

- 説明パネルの中心座標が展示物よりも下に位置する
- 展示物とパネルの中心座標間のユークリッド距離が最小

本アルゴリズムは,一般的に説明パネルが展示物の下部かつ近くに設置され,どの展示物に対応した説明パネルかが明確になるような配置関係にあるという仮定に基づいている.そのため,説明パネルが展示物の上部に設置されている場合や,展示物に対して複数のパネルによる説明が提供されている場合に,正しく関係を推定することができない.

本アルゴリズムは 1 つの展示物と画像内全ての説明パネルの位置関係に基づいて,その展示物と関係があると思われる説明パネルを 1 つ決定する.そのため,画像内に N 個の展示物が検出された場合,同様の推定を N 回実行することで画像内の全ての展示物と説明パネル間の関係性推定を実施することができる.

3.2 1 対 1 ニューラルネットワーク（1 対 1 モデル）

1 対 1 モデルは,それぞれ 1 つの展示物と説明パネルのバウンディングボックス情報をペアで入力し,それらの間の関係の有無を二値分類する,2 層の全結合層から構成されるニューラルネットワークである.本手法では,損失関数としてクロスエントロピー損失を用いることで, 分類境界の最適化をネットワークの学習に委ねる設計となっており,バイナリークロスエントロピー損失で必要となる閾値の手動設定を必要としない.

本モデルの概要を示した図を図 3 に示す.

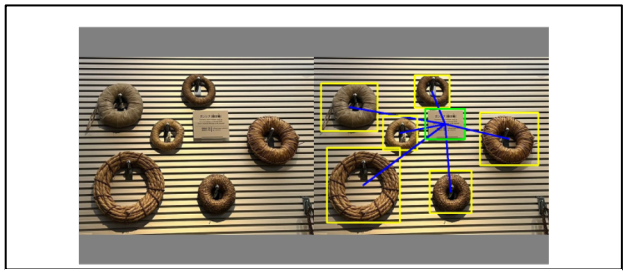

図 2　関係性推定適用例

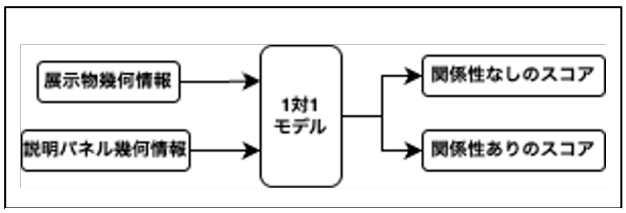

図 3　1 対 1 モデルの概要図

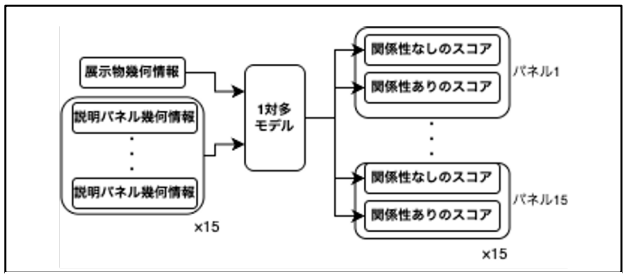

図 4　1 対多モデルの概要図

本モデルは 1 組の展示物と説明パネルの位置情報に基づいて関係性を推定するため,画像内の展示物と説明パネルについて 1 組ずつ独立して関係の有無を推論する.したがって,1 つの展示物に対して複数の説明パネルが設置されている場合であっても対応することができる.また,あらゆる位置関係に対応して関係の有無を推定することができるため,展示物の上部に設置されている説明パネルにも対応することができる.

本モデルは 1 組の展示物と説明パネルの関係性を推定するモデルであるため,画像内に N 個の展示物と M 個の説明パネルが検出された場合,N*M 回の推定を実行して画像内の全ての展示物とパネル間の関係性を推定する.

3.3 1 対多ニューラルネットワーク（1 対多モデル）

1 対多モデルは,1 つの展示物と画像内全ての説明パネルを入力として受け取り,それぞれのパネル間の関係の有無を推定する,2 層の全結合層で構成されたニューラルネットワークである.本モデルでは,展示物と複数パネル間の位置関係や全体的な文脈を踏まえた推論が可能となる.

一方で,画像内で検出される説明パネル数は画像ごとに異なるが,1 対多モデルの入力は固定長である.そこで,モデルに入力する説明パネルの数を 15 個に設定し,画像内の説明パネル数が 15 個を超える場合は,展示物と各説明パネル間の距離が近い順に 15 個の説明パネルを選択してモデルに入力し,モデルへの入力に選択されない説明パネルについては関係性無しとして出力する.逆に画像内の説明パネル数が

15 個未満の場合には,余った入力分をゼロパディングで埋めてモデルに入力している.その際,ゼロパディングを行なった入力に対応したマスクを適用することでゼロパディングを行なった入力に関する誤差の逆伝搬を防いで学習に影響を与えないようにして学習を行った.

本モデルにおいても 1 対 1 モデルと同様にクロスエントロピー損失を用いた学習を行っている.そのため,本モデルは 15 個のパネルについて,パネルごとに 2 値分類を行い 15 個のパネルの展示物に対する関係性を予測しており,モデルの出力はパネル数 15*クラス数 2 のベクトルとなる.

本モデルは,15 個を上限とする画像内全ての説明パネルを入力することができるため,1 組の展示物と説明パネルの位置関係だけでなく説明パネル間の位置関係を考慮した上で各説明パネルについて展示物との関係性を推定することができる.1 対多モデルの概要を示した図を図 4 に示す.

本モデルは,1 つの展示物と全ての説明パネルの関係性を推定するモデルであるため,画像内に N 個の展示物と M 個の説明パネルが検出された場合,N 回の推定を実行して画像内の全ての展示物と説明パネル間の関係性を推定する.

4. 実験

4.1 評価手法

本実験の評価手法は,1 組の展示物と説明パネル間の 1 対 1 の関係性に対する予測精度によって判断する.評価指標として,正解率(Accuracy),適合率(Precision),再現率(Recall),F1 値(F1 Score)を用いてモデルの精度や特性を評価する.

本モデルは博物館デジタルアーカイブの自動化に貢献することを目的としている.実際に組み合わせ推定を自動化に組み込む場合には,博物館職員による確認が必要である.この時,実際には関係性がある組み合わせに対して関係性なしと推測している場合,職員が手動で関係性のあるペアを指定し直す必要がある.逆に関係性のない組み合わせに対して関係性があると推定した場合は,その組み合わせを破棄するだけで良いので職員の作業負担が小さくなる.そのため,本実験で構築するモデルは組み合わせありのペアに対する見落としの少なさを評価する再現率を重視してモデルの評価をおこなう.

4.2 実験条件

4.2.1 データセット

本実験で使用するデータセットは,独自に作成したデータを使用した.データセットの各サンプルは画像情報,画像内の展示物と説明パネルのバウンディングボックス情報,関係の有無を示すラベルによって構成されている.

本実験で使用したデータセットの概要を表 1 に示す.

表 1 においてクラス 0 は関係性なし,クラス 1 は関係性ありを示している.

4.2.2 モデルの入力

本実験で作成した 1 対 1 モデルと 1 対多モデルの入力として用いた特徴量は,どちらも展示物と説明パネルのバウンディングボックスに関する幾何情報を入力としており,共通している.使用した特徴量を以下に示す.

- 中心座標（x,y）
- 横幅
- 高さ
- 面積

バウンディングボックスの幾何情報は画像の横幅と高さを用いて正規化してモデルに入力される.そのため,各展示物や説明パネルの特徴量は 5 次元で構成される.

表 1 実験で使用するデータセットの概要

	Train		Validation		Test	
データ数	4992		1248		836	
クラス	0	1	0	1	0	1
データ数	3903	1089	954	294	567	269

4.3 実験結果

4.3.1 ルールベースモデル

ルールベースモデルのテストデータに対する各評価指標の精度を表 2 に示す.

表 2 ルールベースモデルのテストデータに対する精度

	Precision	Recall	F1 Score
関係性なし	0.90	0.95	0.92
関係性あり	0.87	0.77	0.82
正解率		0.89	

4.3.2 1 対 1 モデル

1 対 1 モデルは表 1 に示した学習データのクラス不均衡への対処として各クラスに重みを付けて学習を行なった.クラスごとの重みは,学習データ数を各クラスのデータ数で割って算出した.学習時に使用した重みを以下に示す.

- クラス 0（関係性なし）：1.28
- クラス 1（関係性あり）：4.58

1 対 1 モデルのテストデータに対する各評価指標の精度を表 3 に示す.

表 3 1 対 1 モデルのテストデータに対する精度

	Precision	Recall	F1 Score
関係性無し	0.93	0.82	0.87
関係性あり	0.70	0.87	0.78
正解率		0.84	

4.3.3 1 対多モデル

1 対多モデルも 1 対 1 モデルと同様に学習データセットのクラス不均衡への対応として各クラスのデータ数に応じた重みを適用した上で学習を行なった.使用した重みは 1 対 1 モデルと同様の数値で学習した.

1 対多モデルのテストデータに対する各評価指標の精度を表 4 に示す.

表 4　1 対多モデルのテストデータに対する精度

	Precision	Recall	F1 Score
関係性なし	0.96	0.91	0.93
関係性あり	0.83	0.93	0.87
正解率		0.91	

5. 考察

3 つのモデルを比較すると,ルールベースモデルは「関係性無し」に対する F 値が 0.92 と 1 対多モデル(0.93)とほぼ

同等の値を示した.一方で,「関係性あり」に対する再現率が 3 つのモデル中最も低い結果となった.これは,本実験で採用したアルゴリズムが展示物の上部に位置する説明パネルに対応していない点や,展示物に対して 2 つ以上の説明パネルを紐づけできない点に起因する.その結果,多くの「関係性あり」説明パネルを見逃し,False Negative が増加した.ルールベースモデルは,全体的な正解率が高いモデルであるが,y 座標と距離のみに基づいた判定では,展示物周辺への多方向配置や複数パネル対応,パネルのサイズに応じた相対的な距離関係など,実際の展示環境の多様性への対応に限界がある点が課題である.

1 対 1 モデルは「関係性なし」の適合率と「関係性あり」の再現率を除いてルールベースモデルを下回る精度となった.また,1 対多モデルに対しては全ての精度で下回っている.本モデルは,ルールベースモデルとは異なり 1 つの展示物に対して複数の説明パネルを関連付けられる設計により,「関係性あり」と予測するケースが増加したことで「関係性あり」に対する再現率及び「関係性なし」の適合率がルールベースモデルを上回った.一方で,False Positive の増加により正解率や F 値の精度はルールベースモデルを下回った..本モデルは,展示物と説明パネルのペアごとに独立して推論を行う構造であるため,他のパネルとの相対的な位置関係など,画像全体の文脈情報を活用できないという制約がある.そのため,誤検出の抑制に限界がある.

1 対多モデルの精度は,「関係性なし」の再現率と「関係性あり」の適合率を除いて他のモデルを上回る精度を示した.特に,「関係性あり」に対する再現率は他のモデルに対してそれぞれ 16 ポイントと 6 ポイント高い精度を示し,本研究において重視する指標である「関係性あり」の検出性能において顕著な改善を示した.1 対多モデルは 1 つの展示物に対して最大 15 個までのパネル情報を同時に処理可能であり,説明パネル間の相対的な位置関係を踏まえた推論が可能である.複数の説明パネル情報の活用により,高精度な推定を実現した.

以上の結果から,本研究における展示物と説明パネルの関係推定において,各モデルは異なる特性を持つことが明らかになった.ルールベースモデルは高い全体精度を示すものの,固定的なルールによる制約から複雑な位置関係への対応に限界がある.1 対 1 モデルは「関係性あり」の再現率は向上するが,文脈情報の欠如により誤検出が増加し,全体的な精度が低下する.一方,1 対多モデルは複数パネル間の位置関係を考慮することで,本研究で最も重要な「関係性あり」の検出において最も高い性能を達成した.

これらの結果は,博物館における展示物と説明パネルの関係推定において,単純なルールベースや独立したペア判定では限界があり,複数の説明パネル間の文脈情報を統合的に処理することの重要性を示している.展示レイアウトの多様性や複雑性を考慮すると,パネル間の位置関係を活用した 1 対多モデルのアプローチが最も実用的であると結論づけられる.

6. おわりに

本稿では,博物館展示物のデジタルアーカイブ自動化に関する研究として,画像内の複数の展示物と説明パネル間の関係の有無を推定する手法について,3 つのモデルを構築して精度を比較評価した.

実験の結果,1 対多モデルが本研究で最も重要な「関係性あり」の検出において最高の性能を示し,複数の説明パネル間の位置関係を統合的に処理することの有効性が確認された.一方で,ルールベースモデルは高い全体精度を示したものの,複雑な位置関係への対応に限界があり,1 対 1 モデルは文脈情報の欠如による誤検出の増加が課題として明らかになった.

今後の課題として,より多様な展示レイアウトに対応するためにデータセットの拡充が必要である.

謝辞

本研究は JSPS 科研費 JP22K01014 の助成を受けたものです.

参考文献

[1] みずほ総合研究所株式会社, “令和 2 年度「博物館ネットワークによる未来へのレガシー継承・発信事業」における「博物館の機能強化に関する調査」”, 2021 年 3 月.

[2] 赤嶺 有平, 上原 和樹, 根路銘 もえ子, 當間 愛晃, “RODM：普及型モバイルデバイスの LiDAR センサを活用した自律走行ロボット開発フレームワーク”, 情報処理学会論文誌, Vol.65, No.12, pp.1916 – 1925, (2024)

[3] Danfei Xu, Yuke Zhu, Christopher B.Choy, Li Fei-Fei. Scene Graph Generation by Iterative Message Passing. CVPR. Pp.5410-5419.(2017)

[4] R.Krishna, Y.Zhu, O.Groth, J.Johnson, K.Hata, J.Kravitz, S.Chen, Y.Kalantidis, L.-J.Li, D.A.Shamma, M.bernstein, and L.Fei-Fei, Visual genome: Connecting language and vision using crowdsourced dense image annotations. International Journal of Computer Vision (IJCV), Vol.123, pp.32-73, (2017)