

鉄製遺物の劣化特性に基づく分類に向けた 画像特徴量の統計的検討

山田圭吾¹ 福永香¹ 柳田明進² 脇谷草一郎²

概要：金属製遺物の収蔵・展示には、考古学的分類に加え、劣化状態の把握が不可欠である。本研究では、自動劣化診断に向けた基礎的検討として、平城宮跡出土の金属製遺物画像を対象に、YOLO および SAM2 を用いた対象領域の抽出およびデータセットの構築を行った。得られた画像に対して GLCM によるテクスチャ解析を適用した結果、赤金鉱の有無に応じて特徴量に差異が確認され、GLCM 特徴量が劣化状態の識別に有効な指標となる可能性が示唆された。

キーワード：金属製考古資料、劣化状態、画像セグメンテーション、GLCM 特徴量

Statistical study of image features for classification based on deterioration characteristics of iron artifacts

KEIGO YAMADA^{†1} KAORI FUKUNAGA^{†1}
AKINOBU YANAGIDA^{†2} SOICHIRO WAKIYA^{†2}

Abstract : The proper curation and exhibition of excavated metal artifacts require not only archaeological classification but also an assessment of their deterioration state. As a fundamental step toward automated deterioration diagnosis, this study constructed an image analysis dataset for metal artifacts excavated from the Heijo Palace Site, using YOLO and SAM2 for target region extraction and dataset construction. Texture analysis based on the Gray-Level Co-occurrence Matrix (GLCM) was then applied to the extracted images. The results revealed differences in GLCM features depending on the presence or absence of akaganeite, suggesting that these features have the potential to serve as effective indicators for identifying the deterioration state of metal artifacts.

Keywords : metal archaeological artifacts, deterioration state, GLCM features, image segmentation

1. はじめに

出土品の適切な収蔵・展示環境を構築するためには、考古学的分類に加え、劣化状態の把握が不可欠である。適切な診断を伴わない保管は腐食の進行を招くおそれがある一方、膨大な資料すべてを専門家が評価することは現実的ではない。近年では、考古学分野における機械学習の活用が注目されており[1]、金属製遺物への応用も期待されている。日常的に記録される画像から劣化状態を判別できれば、保存修復や保管条件の検討に寄与する。鉄製遺物では、展示・収蔵中に進行する劣化の程度が埋蔵環境で腐食層内部に形成される塩化物塩の有無に大きく依存することが知られており、塩化物塩を含む遺物では腐食生成物である赤金鉱が発掘後に形成される[2]。そこで本研究では、自動劣化診断に向けた基礎的検討として、平城宮跡出土の金属製遺物画像[3]を対象に、画像解析用データセットを構築した。そして、赤金鉱を含む2件、含まない18件の画像から GLCM 特徴量を算出し、識別可能性を初期的に検討した。

2. データセット構築

金属製遺物画像には、撮影時に記録された定規や管理番号などが含まれており、これらは資料整理には有用である一方、画像解析や機械学習においては対象物の認識を妨げる外来ノイズとなる。図1に、元画像およびその画像を Vision Transformer[4]に入力して得られた Attention Map を示す。図1より、モデルの注意が遺物だけでなく、周囲の定規や記録用の文字・数字にも向けられていることが確認できる。このことから、背景情報が特徴抽出に影響を及ぼし、色調解析やエッジ解析などの画像解析においてもノイズとなることが示唆される。そこで本研究では、YOLO および SAM2 を用いて画像中の金属製遺物領域を抽出し、遺物のみを含む画像へトリミングすることで、データセットを構築した。

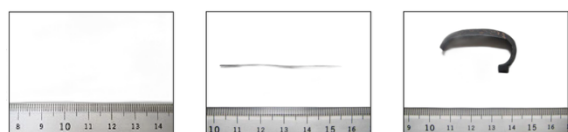
1 横浜国立大学
Yokohama National University
2 奈良文化財研究所
Nara National Research Institute for Cultural Properties



図 1 元画像と Attention Map



(a) 物体検出モデル



(b) 背景除去モデル

図 2 既存モデルをそのまま適用した場合の出力

2.1 YOLO11n

金属製遺物は種類が膨大であり、その全種を学習させることは困難である。また、既存の物体検出・背景除去モデルを直接適用しても、記録写真から遺物のみを取り出すことはできない(図 2)。COCO で事前学習された YOLO11n[5] を記録画像 40 枚に適用したところ、38%の画像で何も検出されず、残りでは airplane, scissors, bird 等の無関係なクラスとして検出された(図 2 (a))。学習データに金属製遺物のクラスが存在しないため、原理的に正しく認識できない。一方、顕著性に基づく背景除去モデル U²-Net[6] では遺物自体は多くの画像で抽出されるものの、40 枚中 7 枚(18%)で併写された定規が前景として残存した(図 2 (b))。特に細い銅針を撮影した図 2 (b) 左端では、高コントラストな目盛りを持つ定規が遺物より顕著と判断され、遺物がほとんど抽出されていない。そこで、遺物自体を 1 クラスとして記録画像 50 枚を YOLO11n に学習させたところ、検証時の mAP50 は最良でも 0.642 に留まり、4 エポック以降は過学習により低下した。この精度では、当初想定していた検出矩形を SAM2 に与える構成において約 3 割の矩形が誤ることとなり、データセット構築に必要な信頼性を満たさない。これは、遺物が形状・色調・大きさにおいて分散が大きいためであると考えられる。そこで本研究では、対象物そのものではなく対象物ではないもの、すなわち定規・文字・管理番号を検出対象とした。この方針には二つの利点がある。第一に、膨大な遺物の全種類を手動でラベリングし学習させ続けることは現実的でない。第二に、発掘調査で収集された遺物は撮影時に定規や管理番号を併写することが慣習であり、形状の異なる遺物であっても記録画像



図 3 ラベリングした箇所



図 4 格子模様を定規と誤検出する例

には定規・文字・番号が共通して写り込む。定規・文字・管理番号は撮影慣習により形状と位置が定型的であり、遺物と異なりクラス内分散が小さいため、少数の学習データでも安定した検出器を構築できる。物体検出には YOLO11n を用いた。YOLO11 の最小構成でありパラメータ数が約 2.6M と軽量なため、少数データでの学習および非 GPU 環境での実行に適する。COCO 事前学習済み重みを初期値とし、記録画像 50 枚に対して手動でラベリングした「定規」「調査次数や資料名称等の文字」「管理番号」の 3 クラスを学習させた。学習条件は入力解像度 640×640、エポック数 120、early stopping patience 40 とし、Apple Silicon の MPS 上で実行した。推論時の信頼度閾値は 0.35 とした。ラベリングした箇所を図 3 に示す。しかし、この検出器の適用過程で、定規クラスに関して新たな問題が確認された。緑方眼紙を背景とする画像において、方眼の格子模様を定規と誤検出する事例である(図 4)。方眼の格子は定規の目盛りとの外見上の類似性が高いため、両者が混同されたと考えられる。そこで、実際の定規が備える幾何学的性質を制約として導入した。今回使用したデータセットでは、定規はいずれも画像幅のほぼ全体を占める帯として画面下部に位置していた。この規則性を利用し、「幅が画像幅の 60%以上」かつ「矩形の下辺が画像下端から 30%以内」の二条件を満たす検出のみを定規として採用した。以上により検出された定規・文字・管理番号の領域を、後段の SAM2 における負例プロンプト、最終マスクからの除外、およびトリミング枠の回避に用いた。

2.2 前景推定

2.1 で検出したラベル領域は対象物ではない位置を与えるが、SAM2 に分割対象を指示するためには、逆に対象物である位置を推定する必要がある。本研究では、記録画像が均質な台紙上で撮影されるという性質を利用し、背景色からの色差に基づいて対象物候補領域を推定した。まず、画像の四隅および上辺中央の計 5 箇所から矩形パッチを取得し、その画素値の中央値を背景色の基準とした。左右辺の中央は横長の遺物が接触しうるため、また下辺の中央は定規が写り込むため、参照位置から除外した。上辺中央を加えたのは、画像上部のみが明るいといった照明勾配に対応するためである。次に、各画素について 5 つの基準色との最小ユークリッド距離を算出し、大津の二値化法[7]により閾値を自動決定して、背景色から遠い画素を対象物候補とした。閾値を画像ごとに定めるのは、撮影条件による背景の明るさの変動を吸収するためである。さらに、背景まで候補に含まれる場合に備え、候補領域が画像面積の 40% を超える場合には閾値を 1.4 倍して再計算する処理を最大 6 回繰り返した。得られた候補の例を図 5 に示す。ただし、色差のみでは検出できない対象もある。銅製遺物の緑青は淡い青緑色で台紙との色差が小さく、閾値処理で除外されやすい。鍍金や真鍮の光沢部も彩度が低く背景と区別しにくい。そこで、台紙が緑や青に着色されることはないという前提のもと、背景色が緑系でない場合に限り、緑～青の有彩画素と、黄金色域の画素を無条件に候補へ加えた。

最後に、候補領域から 2.1 で検出したラベル領域を除いて連結成分解析を行い、画像中央に近く面積の大きい成分を対象物として、その重心を正例点とした。鉄環のような中空の遺物では重心が空洞に落ちるため、その場合は最寄りの成分画素に移動させた。

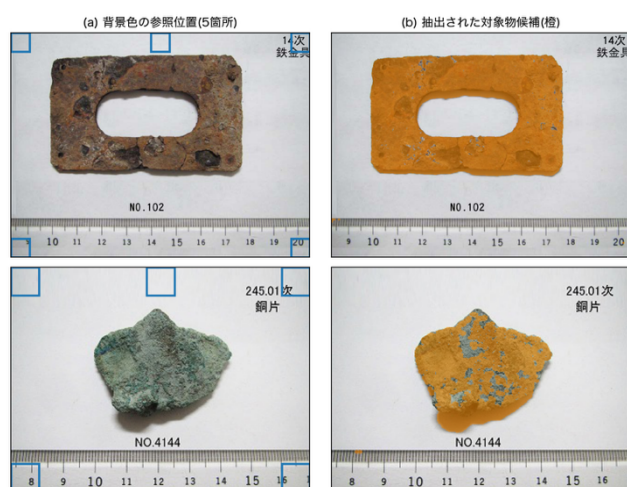


図 5 背景色からの色差による対象物候補の抽出

2.3 SAM2

SAM2 (Segment Anything Model 2) [8] は、点や矩形などのプロンプトで指示した領域を分割する汎用セグメンテーションモデルであり、クラス概念を持たず物体の境界のみを学習している点に特徴がある。そのため、学習データに存在しない金属製遺物に対しても追加学習なしに輪郭を抽出できる。本研究では sam2.1_s を事前学習済みのまま使用した。SAM2 は指示された位置の物体を分割するため、何を対象とするかはプロンプトの設計に依存する。本研究では、2.2 で決定した対象物上の点を正例点、画像四隅および 2.1 で検出したラベル領域の中心を負例点として与えた。四隅の負例は台紙を、ラベル領域の負例は定規・文字・管理番号を、それぞれ対象から明示的に排除するために置いている。ただし、ラベル領域の負例は 2.1 の YOLO 検出器の出力に依存するため、誤検出があると SAM2 の分割を誤って制約しうる。そこで、プロンプトは (1) 正例点+四隅+ラベル領域、(2) 正例点+四隅、(3) 正例点のみ、の 3 段階を用意し、制約の少ない段階へ段階的に単純化できるようにした。SAM2 は候補マスクごとに、正解と照合せずモデル自身が算出する予測 IoU を出力し、各段階でこの予測 IoU が 0.7 以上のマスクが得られた時点で、それ以降の段階は試さずに採用する。プロンプトの配置と得られたマスクの例を図 6 に示す。マスク面積が画像の 35% を超える、または 2.2 で得た対象物候補画素のうちマスクに含まれる割合が 50% 未満の場合は、台紙全体を誤って分割したとみなす。この場合、画像を 3×3 に区切った格子点 9 箇所、および対象物候補の連結成分の重心を 1 点ずつ単独で試し、条件を満たすもののうち最も信頼度が高いマスクを採用する。1 枚に複数遺物が写る場合があるため、マスクに含まれない対象物候補の連結成分が残っていれば、それぞれに正例点を与えて再分割し、マスクへ統合する。以上により、SAM2 は 1 枚あたり複数回呼び出される。大半の画像は初回のプロンプトで完了し、困難な画像にのみ計算資源を集中させる適応的な構成である。

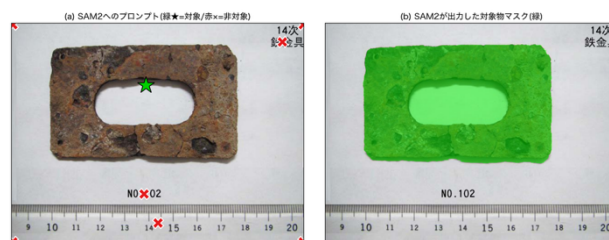


図 6 SAM2 のプロンプト設計と分割結果

2.4 トリミング

得られた対象物マスクをもとに、遺物のみを含む矩形へトリミングを行った。まず、マスクからノイズとなる小さな成分を除いた主要部分の外接矩形を、トリミング枠の基準とする。しかしこの基準だけでは、マスクが取りこぼした遺物の破片が枠から漏れることがある。そこで、2.2と同じ色差に基づく候補検出を用いて、基準の外接矩形の近傍にある対象物候補領域を別途抽出し、これを基準の外接矩形と統合した。この統合後の矩形に、各辺6%の余白を加えたものをトリミング枠の初期候補とした。次に、この枠が2.1で検出したラベル領域と重なる場合、重なりが解消されるまで該当する辺を内側へ縮小した。さらに、余白部分に文字のインクが残る場合には、インクの直前まで最小限の縮小を行った。一方、これらの縮小によって遺物が枠境界で切断される場合があるため、マスクが枠境界に接する辺は、接触が解消されるまで外側へ拡張した（ただし画像端およびラベル領域は越えない）。この基準で問題が解消しない場合は、近傍候補や余白を段階的に削って再試行し、最も問題の少ない枠を採用する。トリミング枠を図7に示す。黄色の実線が最終的なトリミング枠、青の破線が対象物マスクの外接矩形、赤の破線が検出したラベル領域（定規・文字・番号）を表す。

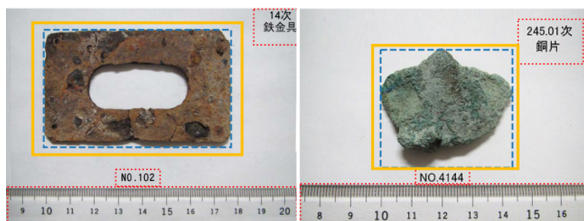


図7 トリミング枠

2.5 構築したデータセット

以上の処理を、鉄製品 1,944 枚および銅製品 844 枚の計 2,788 枚の記録画像に適用した。対象領域の抽出およびトリミングは、大部分の画像で自動的な品質検証を通過した。学習を要したのは 2.1 のラベル検出器のみであり、その学習データは手動でラベリングした 50 枚である。すなわち、遺物 1 枚ごとの人手によるアノテーションを行うことなくデータセットを構築した。構築したデータセットの例を図8に示す。



図8 構築したデータセットの例

3. 特徴量解析

3.1 マスク付き GLCM 特徴量

構築したデータセットのうち、鉄製遺物に対し、劣化状態の識別可能性を検討するための初期的な解析として、マスク付き GLCM (Gray-Level Co-occurrence Matrix) によるテクスチャ解析を適用した。GLCM は Haralick ら[9]により提案された手法であり、画像中で一定の距離と方向だけ離れた画素対の濃淡値の組み合わせが出現する頻度を行列として表現する。この行列から算出される統計量により、画像の粗さ、均質性、規則性といったテクスチャの性質を定量化できる。考古資料の表面分析への適用例もあり、Sferrazza[10]はフリント石器の使用痕分類において、GLCM 特徴量が表面性状の記述に有効であることを示している。共起行列の作成例を図9に示す。本研究で GLCM を採用したのは、標本数の制約による。赤金鉱を含む資料は 2 件、含まない資料は 18 件と少なく、教師あり学習による分類器を構築しても、学習後の検証は事実上困難である。自己教師あり学習であれば約 1,900 件の未分析画像を用いた表現学習自体は可能だが、最終的な検証段階では同じ制約が残る。深層学習の枠組みを用いても標本数の問題自体は解決しないことから、学習を要さず算出できる統計的特徴量である GLCM を用いた検討を行った。得られた結果は、赤金鉱の確定的な識別基準としてではなく、今後、未分析資料の中から機器分析等による確認対象を優先的に選定するための一指標の候補として位置づけられる。前景の画素値を 1 から 32 の 32 階調に量子化し、背景を 0 とした上で値 0 を含む画素対を共起行列から除外した。これにより、白色の台紙と遺物との境界に生じる強いコントラストが特徴量に混入することを防いだ。この処理は、2 章で構築した対象物マスクがあって初めて可能となる。画素対の距離は 1 とし、 0° 、 45° 、 90° 、 135° の 4 方向について 13 種の Haralick 特徴量を算出した。算出には mahotas ライブラリ[11]の haralick 関数を用いた。方向依存性を除くため 4 方向の平均を、方向による変動を捉えるため 4 方向のレンジ（最大値と最小値の差）を、それぞれ特徴量とし、計 26 次元とした。

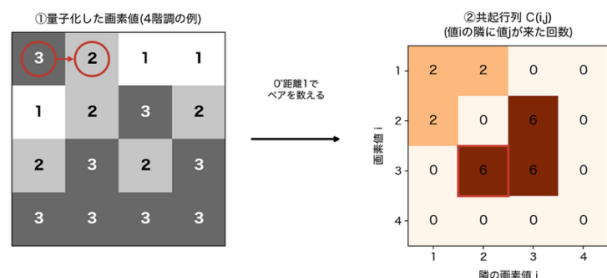


図9 共起行列の作成例

3.2 統計的評価

赤金鉱を含む2件と含まない18件の間で、各特徴量に差があるかを評価した。標本数が極端に少なく正規性も仮定できないため、分布に仮定を置かない順列検定を用いた。順列検定は、全ての組み合わせを列挙できる場合には近似ではなく厳密なp値を与える[12]。本研究では20件から2件を選ぶ配置が $C(20, 2) = 190$ 通りに限られるため、その全通りについて効果量を算出して帰無分布を構成し、真のラベル配置の極端さを両側で評価した。ただし、この構成ではp値の下限が $2/190 \approx 0.0105$ となり、完全な分離が観測されてもこれを下回ることはない。したがって、下限に達した特徴量間の比較には効果量を用いる必要がある。効果量にはCliff's delta[13]を用いた。これは2群間の順位の重なりを-1から+1の範囲で表すノンパラメトリックな指標であり[14]、絶対値が1のとき2群は完全に分離し、0のとき完全に重なる。分布の形状を仮定しないため、標本数が少なく分布が未知である本研究のデータに適し、複数の特徴量にわたる客観的な比較・順位付けを可能にする。

表 1 全特徴量の結果(p<0.05で有意)

特徴量	Cliff's δ	p値	有意差	特徴量	Cliff's δ	p値	有意差
Contrast (平均)	+1.000	0.0105	●	Sum Entropy (レンジ)	+0.722	0.1263	
Contrast (レンジ)	+1.000	0.0105	●	ASM (レンジ)	-0.722	0.1263	
Difference Entropy (平均)	+1.000	0.0105	●	Difference Entropy (レンジ)	+0.667	0.1684	
Difference Variance (平均)	-1.000	0.0105	●	Sum Entropy (平均)	+0.500	0.3158	
Sum Variance (レンジ)	+1.000	0.0105	●	Difference Variance (レンジ)	-0.500	0.3158	
Correlation (平均)	-0.944	0.0211	●	Variance (平均)	+0.444	0.3789	
Correlation (レンジ)	+0.944	0.0211	●	Sum Average (平均)	+0.389	0.4421	
IDM (平均)	-0.944	0.0211	●	Sum Variance (平均)	+0.389	0.4421	
Entropy (平均)	+0.944	0.0211	●	IMC1 (レンジ)	-0.278	0.5895	
IMC2 (レンジ)	+0.833	0.0632		Sum Average (レンジ)	+0.278	0.5895	
IMC1 (平均)	+0.833	0.0632		IDM (レンジ)	-0.111	0.8526	
IMC2 (平均)	-0.778	0.0947		Entropy (レンジ)	+0.111	0.8526	
ASM (平均)	-0.778	0.0947		Variance (レンジ)	-0.111	0.8526	

4. 結果

26 種の特徴量に順列検定を適用した結果、9 種において有意差 (p<0.05) が認められた。全特徴量の結果を表 1 に示す。このうち、Contrast の平均およびレンジ、Difference Entropy の平均、Difference Variance の平均、Sum Variance のレンジの5種では、Cliff's deltaの絶対値が1.000となり、赤金鉱を含む2件が含まない18件のいずれとも完全に分離した。これらのp値は0.0105であり、本標本数における理論的下限に達している。有意差が認められた特徴量の符号は一貫しており、赤金鉱を含む群ではContrastおよびEntropyが高く、IDMおよびCorrelationが低い。すなわち、近傍画素間の濃淡差が大きく濃淡分布が無秩序で、局所的な均質性と画素間の線形的な依存が弱いことから、赤金鉱を含む遺物の表面は粗く不均質なテクスチャを持つといえる。一例として、Contrast (平均) の分布を図 10 に示す。赤金鉱を含む2件は他のいずれよりも高い値を示している。この結果に基づき、未分析資料のうちContrastが高い資料を優先的に選定し、機器分析による赤金鉱の確認を依頼する予定である。ただし、本結果には標本に起因する限界がある。赤金鉱を含む2件はいずれも同一の出土層(レキ混砂)から出土しており、他の18件に当該層の資料は含まれない。したがって、観測されたテクスチャ差が赤金鉱の生成に起因するののか、当該層の埋蔵環境に由来する要因に起因するののかは、本標本から統計的に分離できない。もっとも、赤金鉱は埋蔵環境下で塩化物塩が形成された鉄製遺物で発掘後に生成されるため[2]、出土層と赤金鉱の生成は因果的に関連しており、単純な擬似相関とは限らない。今後は、異なる出土層に由来する正例の蓄積と機器分析の結果を通じて、出土層の交絡を分離した実用的なスクリーニング基準への精緻化を目指す。

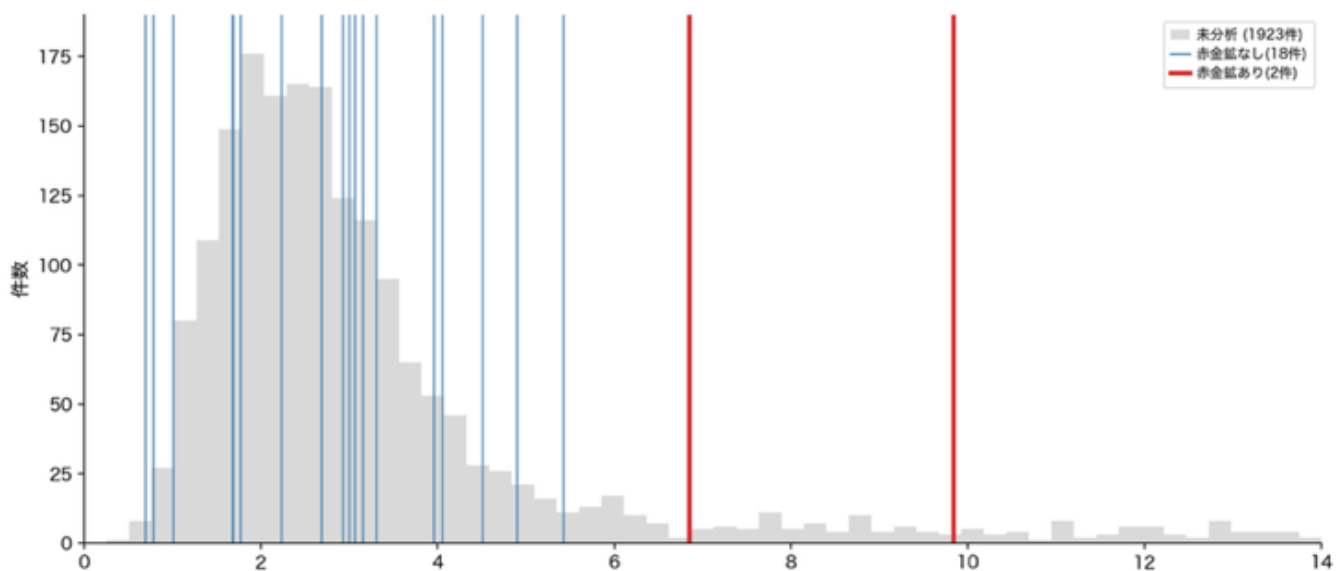


図 10 Contrast (平均) の分布

5. まとめ

本研究では、平城宮跡出土金属製遺物の記録画像 2,788 枚を対象に、YOLO によるラベル検出、背景色差に基づく前景推定、および SAM2 による分割を組み合わせた自動トリミングを適用し、遺物のみを含む画像データセットを構築した。学習を要したのはラベル検出器のみであり、遺物ごとのアノテーションを必要としない。さらに、鉄製遺物 20 件に対するマスク付き GLCM 解析では、赤金鉱を含む 2 件が Contrast をはじめとする複数の特徴量において、含まない 18 件と完全に分離した。標本数の制約と出土層の交絡が残るため、この結果は確定的な識別基準ではないが、多数の未分析資料の中から機器分析による確認対象を優先的に選定するための一指標となりうる。今後は、機器分析による赤金鉱の確認と正例の蓄積を通じて、埋蔵環境の交絡を分離し、スクリーニング基準としての妥当性を検証する。

参考文献

- [1]Bellat, M., Orellana Figueroa, J. D., Reeves, J. S. et al.: Machine Learning Applications in Archaeological Practices: A Review, Journal of Computer Applications in Archaeology, Vol.8, No.1, pp.282-321 (2025).
- [2]Selwyn, L.: Metals and Corrosion: A Handbook for the Conservation Professional, Canadian Conservation Institute (2004).
- [3]柳田明進ほか：土質および埋没過程が鉄製文化財の腐食に及ぼす影響，第 71 回材料と環境討論会要旨集，pp.81-82 (2024).
- [4]Dosovitskiy, A., Beyer, L., Kolesnikov, A. et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, Proc. ICLR (2021).
- [5]Ultralytics: YOLO11 (online), available from <https://docs.ultralytics.com/models/yolo11/> (accessed 2026-07-15).
- [6]Qin, X., Zhang, Z., Huang, C. et al.: U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection, Pattern Recognition, Vol.106, 107404 (2020).
- [7]Otsu, N.: A Threshold Selection Method from Gray-Level Histograms, IEEE Trans. Systems, Man, and Cybernetics, Vol.9, No.1, pp.62-66 (1979).
- [8]Ravi, N., Gabeur, V., Hu, Y.-T. et al.: SAM 2: Segment Anything in Images and Videos, arXiv preprint arXiv:2408.00714 (2024).
- [9]Haralick, R. M., Shanmugam, K. and Dinstein, I.: Textural Features for Image Classification, IEEE Trans. Systems, Man, and Cybernetics, Vol.SMC-3, No.6, pp.610-621 (1973).
- [10]Sferrazza, P.: Grey level co-occurrence matrix and learning algorithms to quantify and classify use-wear on experimental flint tools, Journal of Archaeological Science: Reports, Vol.48, 103869 (2023).
- [11]Coelho, L. P.: Mahotas: Open source software for scriptable computer vision, Journal of Open Research Software, Vol.1, No.1, e3 (2013).
- [12]Ernst, M. D.: Permutation Methods: A Basis for Exact Inference, Statistical Science, Vol.19, No.4, pp.676-685 (2004).
- [13]Cliff, N.: Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions, Psychological Bulletin, Vol.114, No.3, pp.494-509 (1993).
- [14]Meissel, K. and Yao, E. S.: Using Cliff's Delta as a Non-Parametric Effect Size Measure: An Accessible Web App and R Tutorial, Practical Assessment, Research & Evaluation, Vol.29, No.2 (2024).