

GRAIL: A Multimodal Genomic, Visual, and Blockchain Framework for Cultural Heritage Authentication

Giorgio Fornara^{1,*†}, Roberto Pirrone^{1,†}, Stefania Montani^{2,†} and Manuel Striani^{2,†}

¹Digital T  chne Sagl, Lugano, Switzerland

²Computer Science Institute – DiSIT, University of Piemonte Orientale, 15121 Alessandria, Italy

Abstract

Cultural heritage authentication is a physical–digital consistency problem: a digital record is reliable only if it remains bound to the corresponding physical object and to its observable visual state. This paper presents a compact multimodal framework that combines three complementary layers: a synthetic genomic DNA marker used as a biological identity anchor, a blockchain-based dossier for tamper-evident provenance records, and a Vision Transformer model for object-specific visual fingerprinting. Instead of treating authentication as a single binary classification task, the framework evaluates consistency among genomic identity, visual similarity and provenance continuity. We formalise the main verification primitives, define a threat model for object substitution and record manipulation, and outline an experimental protocol based on intra-object stability, inter-object separability, false acceptance/rejection rates, acquisition robustness and longitudinal embedding drift. The contribution is a computer-science-oriented architecture for integrating molecular marking, visual representation learning and trustworthy digital records in cultural heritage authentication and preservation.

Keywords

Cultural heritage authentication, Vision Transformer, Visual fingerprinting, Synthetic DNA marker, Blockchain provenance, Multimodal verification, Artwork verification

1. Introduction

Forgery, uncertain provenance and weak traceability remain major challenges for artworks, collectibles and cultural heritage objects. Expert connoisseurship, archival research, material analysis and conservation records are indispensable, but they are not always scalable, reproducible or digitally persistent. A key open issue is the *physical–digital binding*: a database record, digital twin or blockchain entry is useful only if it can be reliably linked to the specific physical object it represents.

Recent advances in computer vision have shown that deep learning can support visual art analysis, including pattern extraction, attribution, style recognition and authentication [1, 2, 3]. Vision Transformers extend this line of work by modelling relationships among image patches through self-attention, which is useful when visual evidence is distributed across local surface details and global composition [4, 5, 6]. In parallel, DNA-based information technologies provide durable physical markers for object identity [7, 8], while blockchain infrastructures have been explored for cultural heritage provenance and authenticity management [9, 10].

These technologies are often considered separately. Visual Artificial Intelligence (AI) can compare images but cannot guarantee that the imaged object is the certified physical object. Blockchain can preserve records but cannot by itself prevent association with the wrong object. Genomic markers can anchor identity but do not describe the visual state or degradation of the artwork. This paper therefore

4th Italian Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2026, <https://ai4ch.di.unito.it/>), co-located with the 24th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2026). 6–9 October 2026, Perugia, Italy

*Corresponding author.

†These authors contributed equally. Contact: info@digitaltechne.ch.

✉ g.fornara@digitaltechne.ch (G. Fornara); r.pirrone@digitaltechne.ch (R. Pirrone); stefania.montani@uniupo.it (S. Montani); manuel.striani@uniupo.it (M. Striani)

ORCID iD 0000-0002-5992-6735 (S. Montani); 0000-0002-7600-576X (M. Striani)



  2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

models authentication as a cross-layer consistency problem involving physical, visual and documentary evidence.

The paper provides four contributions: (i) a multimodal authentication architecture based on genomic marking, blockchain records and Vision Transformer embeddings; (ii) a compact formalisation of artwork verification as cross-layer consistency; (iii) a visual-fingerprinting workflow for certification, verification and condition monitoring; and (iv) an evaluation protocol suitable for computer science assessment under realistic acquisition variability.

2. Background and Related Work

Deep learning has become increasingly relevant to cultural heritage because large digitised collections and high-resolution imaging enable computational analysis of paintings and drawings. Castellano and Vessio review deep learning approaches for pattern extraction and recognition in visual artworks [1]; Elgammal et al. demonstrate stroke-level analysis for attribution and authentication of drawings [2]; and Ugail et al. show how deep transfer learning can support painting authentication in a Raphael case study [3]. These works indicate that visual features can provide useful evidence, but most approaches focus on attribution, classification or binary authenticity. The present work instead focuses on *object-specific visual fingerprinting*: the goal is to verify whether a future acquisition remains visually consistent with a previously certified object.

In the context of the H2020EU SPICE¹ project (Social cohesion, Participation, and Inclusion through Cultural Engagement), DEGARI 2.0 introduces an explainable affective recommender for cultural heritage that combines the TCL commonsense reasoning framework with an ontological formalisation of Plutchik’s wheel of emotions to produce diversity-seeking recommendations beyond same-emotion matching [11]. The subsequent sensemaking system integrates DEGARI 2.0 into GAMGame (developed in cooperation with the Turin Gallery of Modern Art - Galleria d’Arte Moderna, GAM)², where emotion extraction, TCL-based reasoning and story suggestion are used to group and recommend user-generated museum stories from similar and opposite affective viewpoints [12]. A complementary dataset contribution describes the JSON-LDH data layer produced in the same SPICE/GAMGame context, organising artworks, user profiles and user-generated affective contributions into three linked resources (artworks, `ugcUsers` and `ugcContributions`) for modelling citizen curation, community interaction and social cohesion in museums [13].

The work in [14] propose the SPICE Ontology Network (SON), an ontology network for citizen curation that formalises heterogeneous knowledge about artworks, curatorial practices, users, communities, interaction, narratives, symbolism and emotions. Their work shows how semantic web technologies can support knowledge-level reasoning over cultural-heritage data produced by participatory and interpretive activities, including user-generated content, affective responses and community-oriented representations. This perspective is complementary to the present framework: while **GRAIL - Genomic Record and Artwork Identity Ledger** - focuses on physical, digital authentication through genomic seals, blockchain records and visual fingerprints, ontology networks such as SON can provide the semantic layer needed to contextualise certified objects within richer interpretive, social and curatorial knowledge structures. Taken together, these works extend the technical discussion from object analysis to socio-technical interpretation: genomic seals and Vision Transformer fingerprints can certify and monitor cultural objects, while affective sensemaking and dataset layers can support inclusive access, perspective-taking, user-centred interaction and evidence-based analysis of the certified heritage record.

Transformers were introduced for sequence modelling through self-attention [4], and the Vision Transformer adapted this principle to images by processing fixed-size patches as tokens [5]. Hierarchical models such as Swin Transformer [15], improve scalability through shifted windows and multi-scale representations [6]; DeiT reduces the need for very large supervised datasets through data-efficient training [16]; and self-supervised models such as DINOv2 provide general-purpose visual

¹<https://spice-h2020.eu/>

²<https://www.gamtorino.it/it/>

embeddings that may be useful when labelled cultural heritage data are limited [17]. A recent study on art authentication with Vision Transformers confirms the potential of Transformer models for imitation-focused authentication, while also showing that performance depends on the selected contrast set and architecture [18].

DNA-based information technologies and blockchain provenance address a different part of the authentication problem. DNA-of-things demonstrates that synthetic DNA can be embedded in materials as molecular memory [7], and synthetic DNA has been discussed as an information technology for storage, tagging and molecular signal processing [8]. Blockchain-based cultural heritage studies focus on provenance, tokenisation and rights-management scenarios, while also stressing that ledger entries must still be connected to physical objects [9, 10].

Several more specific research lines bear directly on this problem. DNA-based tagging has moved beyond laboratory demonstrations towards the deployed marking of high-value goods, and recent reviews map the chemistries, readout procedures and open reliability issues that condition any genomic seal applied to artworks [19, 20].

From a security perspective, such a seal behaves as a molecular *physical unclonable function*, since its uniqueness derives from physically irreproducible variation rather than from a stored secret [21], and the same perspective clarifies the role of the visual layer: the question is not which class, style or author an image belongs to, but whether two acquisitions depict the same individual object, which is precisely the instance-level verification setting surveyed for deep instance retrieval [22].

Combining such heterogeneous signals into a single decision, follows established practice in multimodal biometrics, where score- and decision-level fusion yields more robust verification than any single modality [23]. What remains specific to our setting is that the fused evidence must ultimately be written into a distributed ledger, which makes physical-digital binding an instance of the *blockchain oracle problem*: on-chain records inherit the trust assumptions of whatever off-chain process reports the physical state, so a tamper-evident dossier is only as reliable as the genomic and visual attestations recorded in it [24].

Our framework combines these lines of research by using DNA as a physical identity anchor, blockchain as a tamper-evident dossier, and Vision Transformers as visual consistency models.

3. Proposed Authentication Framework

The Digital T  chne framework follows a defence-in-depth design. Each layer provides an independent signal: molecular identity, cryptographic provenance and visual representation learning. The final decision is not a single autonomous classification, but a structured assessment of cross-layer consistency.

Table 1

Three-layer authentication architecture of the Digital T  chne framework.

Layer	Technology	Authentication role
Genomic seal	Synthetic DNA marker, unique for each object	Provides a physical identity anchor independent of digital systems.
Blockchain dossier	On-chain record and digital twin	Stores hashes, provenance events and verification history for tamper-evident auditability.
ViT analysis	Vision Transformer visual-fingerprint model	Compares certification and verification embeddings to detect substitution, alteration or degradation.

At certification time, the object is associated with a synthetic nucleotide sequence, imaged under a controlled acquisition protocol and registered in a digital dossier. The dossier stores cryptographic hashes rather than raw sensitive information, including the genomic marker hash, metadata hash and visual-embedding hash. At verification time, the system checks whether the current genomic marker, visual embedding and provenance state are consistent with the certified record. A valid genomic marker with low visual similarity may indicate damage, restoration or substitution of the image; conversely,

Algorithm 1: Compact verification logic for cross-layer consistency

Input: Current image x_t , genomic marker g_t , dossier $R(o)$

Output: Structured verification status

- 1 $h_g \leftarrow \text{Hash}(g_t)$;
 - 2 $z_t \leftarrow f_\theta(x_t)$;
 - 3 Retrieve h_g, h_z , and provenance state P from $R(o)$;
 - 4 $C_g \leftarrow (h_g = h_g)$;
 - 5 $C_v \leftarrow (\cos(z_0, z_t) \geq \tau)$;
 - 6 $C_p \leftarrow \text{validate}(P)$;
 - 7 **return** status(C_g, C_v, C_p);
-

high visual similarity with an invalid marker may indicate image reuse or physical–digital mismatch. Formally, let an artwork be denoted by o . The certification record is:

$$R(o) = \langle h_g, h_m, h_z, P, T \rangle, \quad (1)$$

where h_g is the hash of the genomic marker, h_m is the metadata hash, h_z is the hash of the reference visual embedding, P is the provenance state and T is the timestamped verification history. Given a reference image x_0 and a Vision Transformer encoder f_θ , the visual fingerprint is $z_0 = f_\theta(x_0)$. At a later time t , the verification image x_t is encoded as $z_t = f_\theta(x_t)$, and visual consistency is estimated through cosine similarity:

$$s_t = \frac{z_0 \cdot z_t}{\|z_0\| \|z_t\|}. \quad (2)$$

The verification is visually consistent when $s_t \geq \tau$, where τ is calibrated on validation data and may depend on artwork type, imaging protocol and expected conservation drift.

4. Vision Transformer Visual Fingerprinting

Artwork authentication is difficult because authentic and forged objects can be globally similar while differing in subtle and spatially distributed features. A visual-fingerprint model should therefore capture both local micro-signatures and global spatial relations. Vision Transformers are suitable for this purpose because self-attention can model interactions between distant patches, complementing convolutional models that are often dominated by local inductive biases [5, 6].

The visual layer focuses on features that are relevant to both authentication and conservation: brush-stroke texture and directionality; craquelure morphology; pigment granularity and surface response; relief and impasto patterns; and global compositional relations. Depending on the acquisition setup, the same workflow can be applied to standard RGB images or extended to multimodal imaging, including raking-light, infrared, ultraviolet or multispectral captures.

The operational workflow is uniform across certification and later inspections. First, the object is imaged and preprocessed using a documented acquisition protocol. Second, the Vision Transformer extracts an embedding that represents the visual state of the object. Third, the embedding is compared with the certified reference using calibrated similarity metrics. Fourth, repeated acquisitions are interpreted as a trajectory of embeddings over time, allowing the system to distinguish expected ageing or documented restoration from abrupt visual drift. The compact cross-layer verification procedure combining visual, genomic and provenance evidence is formalized in Algorithm 1.

The returned status should be interpretable by cultural heritage professionals. Instead of only reporting “authentic” or “fake”, the system should distinguish fully consistent, visually and genomically divergent, provenance-incomplete and uncertain cases. This design supports human-in-the-loop decision making and avoids reducing expert authentication to an opaque machine output.

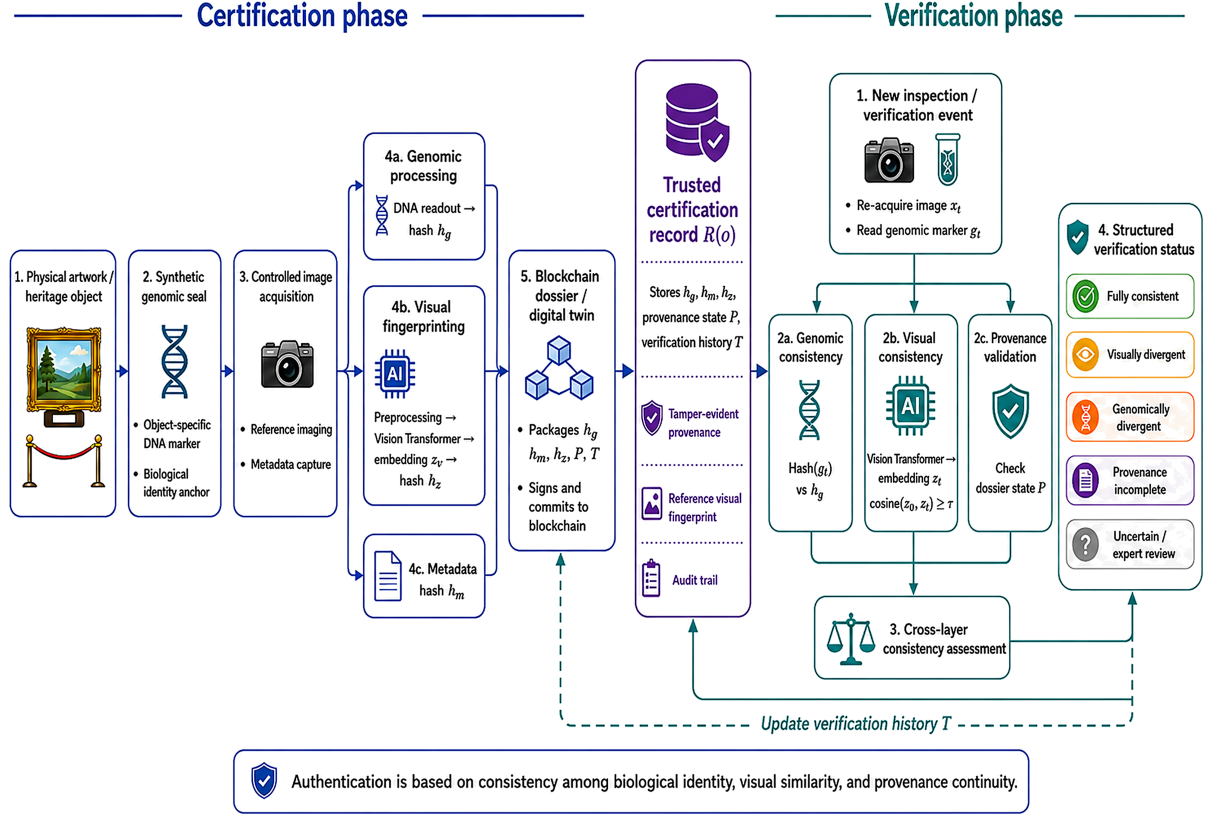


Figure 1: The **GRAIL** pipeline for multimodal cultural heritage authentication based on genomic identity, visual fingerprinting, and provenance continuity.

5. Evaluation Protocol

The Vision Transformer component is designed for real certification and inspection workflows; therefore, its evaluation must consider both machine-learning performance and operational robustness. In the proposed **GRAIL** pipeline, illustrated in Figure 1, authentication is organised as a two-phase process that connects an initial certification event with subsequent verification events through a persistent trusted record. During certification, the physical artwork or heritage object is associated with a synthetic genomic seal, acquired through a controlled imaging protocol, and processed to generate three complementary digital evidences: the genomic marker hash h_g , the metadata hash h_m , and the visual-fingerprint hash h_z . These elements are stored in a blockchain-based dossier, or digital twin, together with the certification record $R(o)$, the provenance state P , and the verification history T . During verification, a new inspection re-acquires the object image x_t and reads the current genomic marker g_t . The system then performs genomic consistency checking, Vision Transformer-based visual consistency estimation, and provenance validation. The final outcome is produced through a cross-layer consistency assessment that distinguishes fully consistent cases from visually divergent, genomically divergent, provenance-incomplete, or uncertain cases requiring expert review.

For each object o_i , the dataset should include a reference acquisition $x_{i,0}$ and multiple verification acquisitions $\{x_{i,1}, \dots, x_{i,n}\}$ collected across different sessions, devices, lighting conditions, or conservation states. This structure reflects the operational behaviour of **GRAIL**: the reference image $x_{i,0}$ is used to compute the certified embedding $z_0 = f_\theta(x_{i,0})$, whereas each later acquisition $x_{i,t}$ produces a verification embedding $z_t = f_\theta(x_{i,t})$. Negative examples should include reproductions, visually similar works, altered versions, and same-artist distractors, since authentication usually requires discrimination among near neighbours rather than unrelated objects. Moreover, the evaluation protocol should include cross-layer mismatch scenarios, such as valid visual similarity with an invalid genomic marker, valid genomic identity with visual divergence, and incomplete provenance records. These cases are necessary to verify

whether the system exploits the multimodal structure of the pipeline rather than relying on a single evidence layer.

The visual encoder should be compared with representative baselines, including ResNet50 and EfficientNet for convolutional transfer learning, ViT-B/16 for standard patch-based transformer modelling, Swin Transformer for hierarchical shifted-window attention, and DINOv2 embeddings for self-supervised visual representation learning [5, 6, 17]. This comparison is relevant because previous art-authentication studies suggest that convolutional and transformer-based models may behave differently depending on the type of imitation and the design of the contrast set [18]. In the GRAIL setting, the comparison should therefore be performed not only at the visual-similarity level, but also at the verification level, by measuring how reliably each encoder supports the cross-layer assessment when combined with genomic and provenance evidence.

The core evaluation metrics should include intra-object stability, inter-object separability, false acceptance rate, false rejection rate, equal error rate, acquisition robustness, and longitudinal drift. Intra-object stability measures the similarity among repeated acquisitions of the same object, whereas inter-object separability evaluates the distance from other objects and near-neighbour distractors. False acceptance and false rejection rates quantify, respectively, the risk of accepting non-matching objects as authentic and rejecting valid objects. Acquisition robustness measures sensitivity to lighting, camera, resolution, cropping, and compression, while longitudinal drift evaluates embedding variations over time in relation to documented conservation events. At the pipeline level, additional indicators should assess the correctness of the final structured status, including the identification of genomic divergence, visual divergence, and provenance incompleteness. This enables the evaluation of GRAIL as an integrated multimodal authentication workflow rather than as an isolated visual-similarity model.

6. Discussion

The framework addresses a recurrent limitation of purely digital provenance systems: a ledger can preserve records, but the record still requires a trustworthy link to a physical object. By combining a genomic marker with a visual fingerprint, the system provides two independent object-binding mechanisms. This may support museum documentation, loan management, certification renewal, insurance assessment and high-value art-market transactions. Periodic visual-fingerprint analysis may also provide data-driven alerts when the visual state of an object changes beyond expected variation.

The current implementation is under development and validation in operational certification scenarios. The genomic identification layer should be applied to selected artworks and cultural objects with institutional partners, including the “*Fondazione Sandretto Re Rebaudengo*”³ and the “*Archivio Gramberg*”⁴. The blockchain infrastructure is operational on the ICP mainnet and supports provenance recording and digital asset management⁵. The Vision Transformer component is the main ongoing research activity in collaboration with the Università del Piemonte Orientale, with forthcoming validation focused on repeated acquisitions, certified artwork datasets and comparison with CNN and transformer baselines.

Several limitations must be considered. Genomic marking must be conservation-compatible, documented and reversible when required. Visual models are sensitive to dataset bias, domain shift and acquisition variability; their outputs should be interpreted as probabilistic evidence rather than absolute truth. Blockchain records improve auditability but do not automatically establish legal ownership or historical authenticity. For these reasons, the system is designed as human-in-the-loop decision support: curators, conservators, art historians and legal experts remain responsible for final interpretation.

³<https://fsrr.org/>

⁴<https://www.biblhertz.it/it/werner-gramberg>

⁵<https://internetcomputer.org/>

7. Conclusions and Future Work

This paper presented **GRAIL**, *Genomic Record and Artwork Identity Ledger*, a compact multimodal framework for cultural heritage authentication based on synthetic genomic marking, blockchain provenance and Vision Transformer visual fingerprinting. The main contribution is to shift authentication from isolated signals to cross-layer consistency analysis. The genomic seal anchors identity to the physical object; the blockchain dossier records provenance and verification history; and the Vision Transformer captures visual evidence about object appearance, alteration and condition.

Future work will focus on large-scale experimental validation, metric-learning optimisation, multi-modal imaging, explainability through attention and saliency analysis, and the integration of conservation metadata into longitudinal embedding-drift models. The long-term objective is not to replace expert judgement, but to provide reproducible computational evidence that strengthens cultural heritage authentication and preservation workflows.

Declaration on Generative AI

During the preparation of this work, the authors used Claude Opus 5 in order to paraphrase, reword, and improve writing style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] G. Castellano, G. Vessio, Deep learning approaches to pattern extraction and recognition in paintings and drawings: An overview, *Neural Computing and Applications* 33 (2021) 12263–12282. URL: <https://doi.org/10.1007/s00521-021-05893-z>. doi:10.1007/s00521-021-05893-z.
- [2] A. Elgammal, Y. Kang, M. Den Leeuw, Picasso, Matisse, or a fake? automated analysis of drawings at the stroke level for attribution and authentication, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018, pp. 42–50. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/11313>. doi:10.1609/aaai.v32i1.11313.
- [3] H. Ugail, D. Stork, H. Edwards, S. Seward, C. Brooke, Deep transfer learning for visual analysis and attribution of paintings by raphael, *Heritage Science* 11 (2023). doi:10.1186/s40494-023-01094-0.
- [4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: *Advances in Neural Information Processing Systems*, volume 30, 2017. URL: <https://arxiv.org/abs/1706.03762>.
- [5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in: *International Conference on Learning Representations*, 2021. URL: <https://arxiv.org/abs/2010.11929>.
- [6] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022. doi:10.1109/ICCV48922.2021.00986.
- [7] J. Koch, S. Gantenbein, K. Masania, W. J. Stark, Y. Erlich, R. N. Grass, A DNA-of-things storage architecture to create materials with embedded memory, *Nature Biotechnology* 38 (2020) 39–43. URL: <https://doi.org/10.1038/s41587-019-0356-z>. doi:10.1038/s41587-019-0356-z.
- [8] L. C. Meiser, B. H. Nguyen, Y.-J. Chen, J. Nivala, K. Strauss, L. Ceze, R. N. Grass, Synthetic DNA applications in information technology, *Nature Communications* 13 (2022) 352. URL: <https://doi.org/10.1038/s41467-021-27846-9>. doi:10.1038/s41467-021-27846-9.
- [9] E. De Vacchio, F. Bifulco, Blockchain in cultural heritage: Insights from literature review, *Sustainability* 14 (2022) 2324. URL: <https://doi.org/10.3390/su14042324>. doi:10.3390/su14042324.

- [10] H. Stublić, M. Bilogrivić, G. Zlodi, Blockchain and NFTs in the cultural heritage domain: A review of current research topics, *Heritage* 6 (2023) 3273–3289. URL: <https://doi.org/10.3390/heritage6040202>. doi:10.3390/heritage6040202.
- [11] A. Lieto, G. L. Pozzato, M. Striani, S. Zoia, R. Damiano, DEGARI 2.0: A diversity-seeking, explainable, and affective art recommender for social inclusion, *Cogn. Syst. Res.* 77 (2023) 1–17. URL: <https://doi.org/10.1016/j.cogsys.2022.10.001>. doi:10.1016/J.COGLYS.2022.10.001.
- [12] A. Lieto, M. Striani, C. Gena, E. Dolza, A. M. Marras, G. L. Pozzato, R. Damiano, A sensemaking system for grouping and suggesting stories from multiple affective viewpoints in museums, *Hum. Comput. Interact.* 39 (2024) 109–143. URL: <https://doi.org/10.1080/07370024.2023.2242355>. doi:10.1080/07370024.2023.2242355.
- [13] M. Striani, B. Díaz-Agudo, G. Jiménez-Díaz, A. Lieto, R. Damiano, Developing a comprehensive dataset for enhancing social inclusion and cohesion through citizen curation in cultural heritage (short paper), in: R. Damiano, S. Ferilli, M. Striani, G. Silvello, B. S. de' Bianchi (Eds.), *Proceedings of the 3rd Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2024) co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024)*, Bolzano, Italy, November 28, 2024, *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 66–76.
- [14] L. Asprino, R. Damiano, M. Daquino, S. D. Giorgis, A. Gangemi, A. Lieto, B. Sartini, M. Striani, An ontology network for citizen curation, *ACM Journal on Computing and Cultural Heritage* 17 (2024) 72:1–72:30. URL: <https://doi.org/10.1145/3704729>. doi:10.1145/3704729.
- [15] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9992–10002. doi:10.1109/ICCV48922.2021.00986.
- [16] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, Training data-efficient image transformers & distillation through attention, in: *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 10347–10357. URL: <https://proceedings.mlr.press/v139/touvron21a.html>.
- [17] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jégou, J. Mairal, P. Labatut, A. Joulin, P. Bojanowski, DINOv2: Learning robust visual features without supervision, 2023. URL: <https://arxiv.org/abs/2304.07193>. arXiv:2304.07193.
- [18] L. Schaerf, C. Popovici, E. Postma, Art authentication with vision transformers, *Neural Computing and Applications* 36 (2024) 12675–12685. doi:10.1007/s00521-023-08864-8.
- [19] A. Kuzdraliński, M. Miśkiewicz, H. Szczerba, W. Mazurczyk, J. Nivala, B. Książkowski, Unlocking the potential of DNA-based tagging: current market solutions and expanding horizons, *Nature Communications* 14 (2023) 6052. URL: <https://doi.org/10.1038/s41467-023-41728-2>. doi:10.1038/s41467-023-41728-2.
- [20] S. A. Sharief, P. Chahal, E. Alocilja, Application of DNA sequences in anti-counterfeiting: Current progress and challenges, *International Journal of Pharmaceutics* 602 (2021) 120580. URL: <https://doi.org/10.1016/j.ijpharm.2021.120580>. doi:10.1016/j.ijpharm.2021.120580.
- [21] Y. Gao, D. C. Ranasinghe, S. F. Al-Sarawi, O. Kavehei, D. Abbott, Emerging physical unclonable functions with nanotechnology, *IEEE Access* 4 (2016) 61–80. URL: <https://doi.org/10.1109/ACCESS.2015.2503432>. doi:10.1109/ACCESS.2015.2503432.
- [22] W. Chen, Y. Liu, W. Wang, E. M. Bakker, T. Georgiou, P. Fieguth, L. Liu, M. S. Lew, Deep learning for instance retrieval: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2023) 7270–7292. doi:10.1109/TPAMI.2022.3218591.
- [23] A. Ross, A. Jain, Information fusion in biometrics, *Pattern Recognition Letters* 24 (2003) 2115–2125. URL: <https://www.sciencedirect.com/science/article/pii/S0167865503000795>. doi:https://doi.org/10.1016/S0167-8655(03)00079-5, audio- and Video-based Biometric Person Authentication (AVBPA 2001).
- [24] G. Caldarelli, Understanding the blockchain oracle problem: A call for action, *Information* 11 (2020). URL: <https://www.mdpi.com/2078-2489/11/11/509>. doi:10.3390/info11110509.