ARTIFICIAL INTELLIGENCE TO FIGHT ILLICIT TRAFFICKING OF CULTURAL PROPERTY

D. Abate1,6*, A. Agapiou1, K. Toubmas1, A. Lampropoulos2, K. Petrides3, R. Pierdicca4, M. Paolanti5, F. Di Stefano3, A. Felicetti3, E. S. Malinverni4, P. Zingaretti6

1The Cyprus Institute, Athalassa Campus, 2121, Nicosia, Cyprus - (d.abate, an.agapiou, k.toubmas)@cyi.ac.cy
2HYPERTECH S.A., 32 Perikleous St, 15232 Athens, Greece - (a.lampropoulos, k.petrides)@hypertech.gr
3Università Politecnica delle Marche, Dipartimento di Ingegneria Civile, Edile e dell’Architettura, 60131 Ancona, Italy – (r.pierdicca, f.distefano, e.s.malinverni)@staff.univpm.it
4University of Macerata, Department of Political Sciences, Communication and International Relations, VRAI Lab, 62100 Macerata, Italy – marina.paolanti@unimc.it
5Università Politecnica delle Marche, Dipartimento di Ingegneria dell’Informazione (DII), VRAI Lab, 60131 Ancona, Italy – (a.felicetti, p.zingaretti)@staff.univpm.it
6Eratosthenes Centre of Excellence, Department of Resilience Society, 3102, Limassol, Cyprus - dante.abate@eratosthenes.org.cy

KEY WORDS: Artificial Intelligence, Social Media Intelligence, Dark Web, Illicit Traffic, Cultural Heritage

ABSTRACT:

The use of artificial intelligence (AI) has the potential to be highly effective in detecting and monitoring illegal trafficking of cultural heritage (CH) goods through image classification techniques, particularly on online marketplaces where the trade of stolen CH objects has become a major global issue. Traditional investigation methods are no longer adequate, but with the assistance of AI, law enforcement agencies and CH organizations can now boost monitoring capabilities to detect, track, and possibly recover stolen objects more efficiently. AI algorithms can indeed analyze images to identify unique features and characteristics that can be used to determine their authenticity and provenance. Additionally, AI can detect patterns and networks of illicit trafficking, and link stolen objects to their places of origin, facilitating the recovery process. In this context, the SIGNIFICANCE project (Stop Illicit Heritage Trafficking with Artificial Intelligence) has been specifically designed to increase the response capabilities of public authorities and police corps against the illicit trafficking of cultural goods perpetrated through internet channels (i.e., social platforms, web, and dark web). By leveraging the power of Deep Learning (DL), AI can help prevent the loss of invaluable cultural artifacts and ensure that they are returned to their rightful owners and places of origin. This paper presents the results reached by the SIGNIFICANCE AI framework on image datasets collected over the web and social media through crawling algorithms.

1. INTRODUCTION

According to Hulkevych (2023), the trafficking of cultural artifacts is more rampant than ever before. While the consequences of cultural property crimes have been acknowledged since the "The Hague Convention for the Protection of Cultural Property in the Event of Armed Conflict" in 1954, the issue has become increasingly urgent in recent decades (Manacorda and Chappell, 2011).

Various factors have contributed to the expansion of this illicit phenomenon. The globalization of the art market has created fresh opportunities for criminals to profit from selling stolen cultural objects (Yates, 2015). Online marketplaces have made it easier to anonymously trade cultural goods without the need for physical storefronts or auction houses.

Political instability and conflicts in certain regions have resulted in a rise in looting and theft of cultural artifacts (Kar and Spanjers, 2017). During times of war or civil unrest, looters and traffickers view valuable cultural assets as a lucrative source of income or a means of money laundering (Davis and Mackenzie, 2014). Countries with compromised security conditions, particularly those that were once the territory of ancient civilizations, are particularly susceptible to these illegal activities.

Furthermore, the growth of illicit trafficking of cultural goods has been fueled by limited awareness and resources dedicated to prevention and combatting such crimes (Polk and Chappell, 2019; Delepierre and Schneider, 2015; Bokova, 2021). The inability to effectively prevent archaeological looting or trafficking is not confined to failed-state environments or specific regions; it is a challenge faced by many EU Member States (Casertano, 2020).

The outbreak of the COVID-19 pandemic has exponentially worsened the situation. While the antiquities market denies the widespread occurrence of illicit trade in cultural goods, law enforcement agencies overwhelmingly acknowledge the problem and identify organized crime involvement at every stage (Votey, 2021).

Despite unprecedented technological innovations benefiting numerous sectors with new tools, equipment, devices, and specialized knowledge, the heritage sector continues to struggle in fully harnessing these advancements to control the loot-
ing and trafficking of cultural assets (Fincham, 2019; Muheidat and Tawalbeh, 2021; Spalazzi et al., 2021).

This paper is organized as follows. Section 2 provides a description of the approaches that were adopted for the prevention of illicit traffic of CH goods. Section 3 describes the proposed DL-based pipeline. In Section 4, an evaluation of the proposed approach is offered, as well as a detailed analysis of each component of our DL-based framework. Finally, in Section 5, conclusions and discussion about future directions for this field of research are drawn.

2. RELATED WORKS

Public authorities and scientific communities involved in the protection and preservation of CH are failing to conserve and persuade others to do the same, resulting in ongoing destruction due to natural disasters and human-made actions. A significant proportion of the loss is attributed to looters who act for commercial reasons and are indirectly financed by private and sometimes public collectors of antiquities. To address this issue, public authorities, law enforcement agencies, and the research community are joining forces to develop technological solutions. Thanks to the very high priority of the topic, the European Commission is currently developing an action plan against trafficking in cultural goods for 2022-25 as part of the EU strategy on organized crime which reflects in several funding opportunities.

2.1 Research Projects

In recent times, various funded projects have been implemented. One such project is RITHMS (Research Intelligence and Technology for Heritage and Market Security), which aims to enhance the operational capacity of Police and Customs/Border Authorities in combating organized and multi-criminal trafficking in cultural goods through research, technological innovation, outreach, and training. Another project, ENIGMA (Endorsing Safeguarding, Protection, and Provenance Management of Cultural Heritage), focuses on protecting cultural goods and artifacts from man-made threats by contributing to research on identification, traceability, and provenance (Patias and Georgiadis, 2023).

2.2 Stolen Object Databases

Several databases of stolen objects are available today, but not all of them are fully accessible to the public and academia. One of the largest databases is LEONARDO, maintained by the Command for the Protection of Cultural Heritage of the Italian Carabinieri, which contains over 1.2 million stolen objects, images, and theft cases. The LEONARDO database is the reference point for Italian law enforcement agencies and allows for a careful analysis of criminal phenomena related to the illicit trafficking of cultural property. The SWOADS Project and the mobile iTPC App, developed by Carabinieri TPC, are improving and expanding the software components of the LEONARDO database in terms of big data, machine learning, and blockchain technology.

The INTERPOL Stolen Works of Art database is the only international database with certified police information on stolen and missing objects of art. It combines descriptions and pictures of more than 52,000 items from 134 countries and allows users to complement their search by uploading a picture of any object of art and checking it with image-matching software. Studies in machine learning have been recently conducted to explore its applications in various domains related to CH, such as historical document processing and analysis of dance techniques.

The National Stolen Art File (NSAF) functions as a centralized database containing information about stolen art and cultural property. In May 1979, the FBI’s Laboratory Division initiated the NSAF during its research phase. This computerized index acts as a repository for data and images of artwork that has been reported as stolen by local law enforcement agencies and FBI field offices, as well as items that have been recovered.

The Lost Art Database is maintained by the German Lost Art Foundation. It documents cultural property that was either demonstrably seized from their owners between 1933 and 1945 as a result of Nazi persecution, or for which such a seizure cannot be ruled out. In addition, the Lost Art Database lists cultural property seized, relocated, or removed during the Second World War.

2.3 Networks

EU CULTNET is an informal network of law enforcement authorities and experts competent in the field of cultural goods. It aims to strengthen coordination between law enforcement and cultural authorities and private organisations, i.e., antique shops, auction houses, online auctions, by identifying and sharing information on criminal networks suspected of being involved in illicit trafficking of stolen cultural goods.

“NETCHER – NETwork and digital platform for Cultural Heritage Enhancing and Rebuilding” implemented a strong trans-sectoral network as well as Recommendations on the fight against looting and trafficking of cultural goods. The participants of the project set up a European network of relevant operators and a Europe-wide charter of good practices to efficiently fight the illicit trafficking of antiquities.

2.4 Recent Technological Advancement

Winterbottom et al. (Winterbottom et al., 2022) used machine learning to track and analyze the sales of antiquities on auction websites, such as eBay. The authors used a large CNN that was already trained on more generic image tasks as a starting point, and then further train it on a specific dataset composed by the digital collection of the Durham Oriental Museum. They achieved a 73% instance classification accuracy on a diverse dataset of images of artefacts with 24,502 images and 4332 unique object instances.

1 https://rithms.eu

2 https://artertimes.fbi.gov

3 https://www.lostart.de/en/start
3. MATERIAL AND METHODS

The growing significance of AI and DL techniques in Cultural Property Protection is partly attributed to their successful application in areas such as image classification. Convolutional Neural Networks (CNNs), for example, have the potential to surpass human capabilities in object recognition and image analysis, especially in terms of processing large volumes of data, which makes them particularly valuable in this field (Walter, 2023). Similar to other neural networks, a CNN consists of an input layer, an output layer, and several hidden layers in between. These layers perform operations that modify the data to learn specific data features. Once the features have been learned across different layers, the CNN architecture proceeds to classification.

In the context of identifying illicit traffic of Cultural Heritage (CH) goods from images, CNNs utilize the extracted features from these images to learn to distinguish different object classes and subsequently classify new images, enabling them to identify suspicious items based on detected patterns. Through training on a diverse and extensive dataset, a CNN can acquire the ability to recognize subtle differences between images that might be challenging for humans to discern.

The SIGNIFICANCE framework (Abate et al., 2022) harnesses AI to analyze and categorize images with the aim of extracting valuable insights about the artwork they depict. It comprises three distinct phases: Data extraction and storage, Deep Learning for CH Image Classification, and Performance Evaluation, each described in detail.

To conduct a comprehensive evaluation of the framework, the "SIGNIFICANCE dataset," which is a collection of CH-related images specifically compiled for this study using web and social media crawling algorithms (as outlined in Subsection 3.3), was employed.

3.1 Data Extraction and Storage

This initial stage is the most complex and critical phase of the entire process. The primary objective is to identify cultural heritage goods using reliable and up-to-date data. To achieve this goal, a web and Instagram crawling algorithm was developed and executed twice daily. The social media crawler leverages the Instaloader library to connect with Instagram and download posts. To collect relevant data, the algorithm searches for specific hashtags associated with different types of artworks as defined by domain experts (e.g., coins, frescoes, etc.). Only information pertinent to the research objective, such as the post image, description, hashtags, user information, and geolocation (if available), is downloaded.

The downloaded images are then stored in an AWS (Amazon Web Services) S3 (Simple Storage Service) bucket, while the accompanying text data is saved in a JSON file (which contains a reference to the image) and stored in MongoDB, a document-oriented database.

The purpose of collecting supplementary data is to analyze CH objects from a temporal and geographical perspective, as it is essential to consider not only the image itself but also the accompanying information. Once this stage is complete, a filtering process is initiated to refine the data further.

3.2 Deep Learning for Cultural Heritage Image Classification

VGG16, developed by Simonyan and Zisserman in 2014, is a remarkably powerful pre-trained convolutional neural network (CNN) widely recognized for its effectiveness in numerous computer vision tasks, with a specific emphasis on image classification. This architecture is composed of a total of 16 layers, featuring 13 convolutional layers, 5 max-pooling layers, and 3 fully connected layers.

When utilizing VGG16, the input image is expected to have a size of 224×224×3, representing its width, height, and the three color channels (red, green, and blue). The network processes this input image through its layers, performing convolutions and pooling operations to extract hierarchical features at different levels of abstraction. The final output of VGG16 is a probability distribution across a predefined set of classes, indicating the network’s confidence scores for each class.

Due to its architecture and the pre-training process, where it is initially trained on a large dataset like ImageNet, VGG16 has demonstrated impressive performance in image classification tasks. It has achieved notable success in accurately recognizing and categorizing objects within images, making it a popular choice among researchers and practitioners in the field of computer vision.

VGG16 can be a powerful tool for CH image classification that can effectively capture the underlying features of the images and achieve high accuracy (Paolanti et al., 2019).

For CH image classification, the VGG16 network was fine-tuned on the SIGNIFICANCE dataset by replacing the last fully connected layer with a new one that has the appropriate number of output classes. To prevent overfitting and enable faster convergence, the pre-trained convolutional layers were frozen during training, while the new fully connected layer was trained using the SIGNIFICANCE dataset until convergence was achieved.
The classification task involved two phases: Multilabel classification and Artworks features. The goal was to classify artefacts and fields for each artwork. The dataset was split into a training set (80%) and a test set (20%), with class balance maintained in both datasets. To achieve this, a deep learning (DL) model was trained using transfer learning from a pre-trained model on ImageNet. The model was fine-tuned using an adaptive optimizer (Adam) with a learning rate of 10e-5, a mini-batch size of 32, and trained for 50 epochs.

### 3.3 SIGNIFICANCE Dataset

The classes that will be identified as the most important to recognize are the following: coins, frescoes, icons, manuscripts, and others. To illustrate the diversity of these classes, Figure 1 showcases a selection of images from the SIGNIFICANCE dataset. Each artwork in the dataset has been labeled according to various features such as location, period, material, and more. Each feature encompasses different classes, thereby providing a comprehensive overview of the dataset.

Table 1 and Table 2 report the data cardinality (number of images), for each class used for the training and testing of the models. The results are for both the phases defined.

Table 1. Artefact feature.

<table>
<thead>
<tr>
<th>artefact</th>
<th>#samples</th>
<th>field</th>
<th>class</th>
<th># samples for class</th>
<th># all samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>total</td>
<td>train</td>
</tr>
<tr>
<td>icons</td>
<td>1678</td>
<td>location</td>
<td>Cyprus</td>
<td>839</td>
<td>671</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other</td>
<td>839</td>
<td>671</td>
</tr>
<tr>
<td>frescoes</td>
<td>718</td>
<td>location</td>
<td>Cyprus</td>
<td>248</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other</td>
<td>222</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td></td>
<td>period</td>
<td>Byzantine</td>
<td>326</td>
<td>261</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other</td>
<td>326</td>
<td>261</td>
</tr>
</tbody>
</table>

Table 2. Artefact classes.

<table>
<thead>
<tr>
<th>Class</th>
<th># samples for class</th>
<th># all samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total</td>
<td>train</td>
</tr>
<tr>
<td>coins</td>
<td>718</td>
<td>574</td>
</tr>
<tr>
<td>frescoes</td>
<td>718</td>
<td>574</td>
</tr>
<tr>
<td>icons</td>
<td>718</td>
<td>574</td>
</tr>
<tr>
<td>manuscripts</td>
<td>718</td>
<td>574</td>
</tr>
</tbody>
</table>

3.4 Performance Evaluation

Following, the standard metrics used to measure the performance of classification methods are described as follow.

Confusion matrix (Table 3) is a specific table layout that allows visualization of the performance of a deep learning algorithm. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class, or vice versa. The name stems from the fact that it makes it easy to see whether the system is confusing two classes (i.e. commonly mislabeling one as another). Support counts the samples belonging to a ground truth class.
Table 3. Confusion Matrix.

True Positive counts the correctly matched samples. For example, it is predicted that an image shows a coin, and it truly is.

False Positive counts the false predicted samples belonging to other ground truth classes. For example, it is predicted that an image shows a coin, but it is not true.

False Negative counts the false predicted samples belonging to the same ground truth class. For example, it is predicted that an image shows another class, but it shows a coin.

Precision explains how many of the correctly predicted cases actually turned out to be positive. Precision is useful in cases where False Positives are a higher concern than False Negatives:

\[
Precision = \frac{TP}{TP + FP}
\]  
(1)

Recall explains how many of the actual positive cases were predicted correctly with the developed model. It is a useful metric in cases where False Negatives are a higher concern than False Positives:

\[
Recall = \frac{TP}{TP + FN}
\]  
(2)

F1 Score gives the harmonic mean between Precision and Recall metrics. It is maximum when Precision is equal to Recall:

\[
F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall}
\]  
(3)

Accuracy simply measures how often the classifier correctly predicts. Accuracy can be defined as the ratio of the number of correct predictions and the total number of predictions:

\[
Accuracy = \frac{TP + TN}{TP + FP + FN + TN}
\]  
(4)

4. RESULTS

The training phase of the experiments was performed using two GPUs in parallel to speed up the process. However, only one GPU was used in the testing phase to ensure fair and accurate evaluation.

Table 4 reports the results of the classification of Icons Location which examples are reported in Figure 2. Table 5 is the results of Frescos Location classification which examples are reported in Figure 3. Table 6 reports the results of Frescos Period which examples are reported in Figure 4. The results of the Multilabel classification are reported in Table 7.

<table>
<thead>
<tr>
<th></th>
<th>Target class</th>
<th>Other Classes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target class</td>
<td>TP</td>
<td>FN</td>
<td>P</td>
</tr>
<tr>
<td>Other Classes</td>
<td>FP</td>
<td>TN</td>
<td>N</td>
</tr>
<tr>
<td>Total</td>
<td>P’</td>
<td>N’</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Icons-Location Classification Results.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyprus</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Others</td>
<td>0.99</td>
<td>0.98</td>
<td>0.89</td>
<td>0.93</td>
</tr>
<tr>
<td>Overall</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 5. Frescoes-Location Classification Results.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyprus</td>
<td>0.96</td>
<td>0.99</td>
<td>0.97</td>
<td>0.93</td>
</tr>
<tr>
<td>Italy</td>
<td>0.89</td>
<td>0.98</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Others</td>
<td>0.95</td>
<td>0.82</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>Overall</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 6. Frescoes-period Classification Results.

Figure 2. Images of icons location. Cyprus (left), Other (right).

Figure 3. Images of frescoes location. Cyprus (left), Italy (centre), Other (right).
The results of this study demonstrate that the deep learning model achieved high accuracy in classifying the artifacts' typology and fields of each artwork. Specifically, the model accurately identified and classified each of the five classes of artwork (icons, frescoes, coins, manuscripts, and other) with an overall accuracy of 93%. Additionally, the model extracted useful information about the depicted artwork, such as its location, period, and style.

The findings indicate that Deep Learning (DL) models hold great promise in automatically classifying images of artwork and extracting valuable information related to cultural heritage objects. These results have far-reaching implications across various domains, including art history, cultural heritage preservation, and Cultural Property Protection.

Given the escalating issue of illicit trade in cultural heritage goods, the need to develop novel tools and methodologies for identifying and tracking these objects has become crucial. In this context, the utilization of DL models presents a promising approach to address these objectives. By leveraging the capabilities of DL, it becomes possible to enhance efforts in identifying, cataloging, and monitoring cultural heritage items, ultimately contributing to the protection and preservation of cultural property.

The potential impact of employing DL models in this context extends beyond the realms of academia and research. It holds the potential to inform decision-making processes, aid law enforcement agencies, support cultural institutions, and raise public awareness about the significance of cultural heritage and the need for its safeguarding.
5. CONCLUSIONS AND FUTURE WORKS

In conclusion, this study demonstrates the potential of deep learning algorithms, particularly Convolutional Neural Networks (CNNs), in recognizing and identifying cultural heritage goods from images. By training a CNN on a large and diverse dataset of legit CH objects, it achieved a high level of accuracy in classifying new images as objects of interest for further investigations.

This study shows that the proposed image classification solution is effective, especially when using the most recent AI architectures. The SIGNIFICANCE AI framework might have several potential contributions to Cultural Property Protection:

- **Automated processing**: the framework can automate the identification of artwork images, making it faster and more efficient than manual methods;

- **Artwork classification**: The framework can classify artwork images into different categories (e.g., icons, frescoes, coins, manuscripts, and others), making it easier for investigators to search and organize artwork.

- **Specific feature extraction**: The framework can extract specific features (e.g., location, period, etc.) for each type of artwork, providing useful information for investigators and curators to better understand the artwork and its history in case of theft.

Overall, the framework could provide valuable insights and data for law enforcement agencies, art historians, and cultural heritage organizations working to combat the illicit trade of heritage assets.

### ACKNOWLEDGEMENTS

This Project is funded by the European Union’s Internal Security Fund — Police 101003625 — SIGNIFICANCE — ISFP-2019-AG-OPC.

### REFERENCES


