Egyptian Hieroglyphs Localisation Through Object Detection

Lucia Lombardi, Francesco Mercaldo and Antonella Santone
Department of Medicine and Health Sciences "Vincenzo Tiberio", University of Molise, Campobasso, Italy

Keywords: Archeology, Hieroglyphs, Deep Learning, Object Detection, YOLO.

Abstract: Old Egyptians used Hieroglyphic language to record their findings in medicine, engineering, sciences, achievements, their religious views, beside facts from their daily life. Thus, it is fundamentally important to understand and digitally store these scripts for anyone who wants to understand the Egyptian history and learn more about this great civilization. The interpretation of Egyptian hieroglyphs is areasonably broad and highly complex problem, but have always been fascinating with their stories and the ability to be read in several ways rather than one, which is a challenge in itself to be translated to modern languages. In this paper, we adopt the YOLO 8 model which revolutionized object detection with its one-stage deep learning approach. YOLO is designed to classify images and accurately determine the positions of detected objects within them. Using this DL approach, we were able to significantly reduce the time required to investigate the interpretation of hieroglyphs. To ensure the reproducibility of our results, we opted to utilize a publicly available dataset. All the metrics demonstrate the anticipated patterns: precision, recall, mAP 0.5, and mAP 0.5:0.95 are expected to increase as the number of epochs progresses, indicative of the model effectively learning to detect objects from Egyptian hieroglyphs images.

1 INTRODUCTION

Recent advances in automated sensing enabled by the proliferation of drones, robotics and Light Detection and Ranging (LiDAR), Big Data, and artificial intelligence may fuel the next wave of archaeological discovery (Zhou et al., 2023; Mercaldo et al., 2022; Huang et al., 2024). Despite the concerns related to its use (Barucci and Neri, 2020), there is in AI the amazing power and the perceived hope not only to automate human tasks, but also to improve human understanding. Fields such as archaeology, philology and human sciences are now beginning to be permeated from AI, even though its actual role has still to be fully understood. The interpretation of Egyptian hieroglyphs is a reasonably broad and highly complex problem. Hieroglyphs are usually found in different materials, such as stone, ceramics, wood, and papyrus, among others. Visual identification will be significantly different depending on the material used to elaborate them. This problem is because each material provides a distinct colors and textures. On the other hand, many hieroglyphs can change their horizontal orientation to show a different meaning in what they want to express. This work focuses on this specific aspect, integrating the Egyptological perspective with the application of the latest information technologies. The obtained results explore the data processing methodology and the creation of a functional and usable AI model for Egyptologists. The results represent a contribution to the research on automatic translation of ancient Egyptian. The AI model is able to generate reasonably accurate translations and can be used to facilitate the interpretation of ancient Egyptian texts. However, it also shows that the research that can be done in this direction is vast and requires further study.

2 THE METHOD

This section introduces the proposed method designed to identify and pinpoint Egyptian hieroglyphs within images. Specifically, we present an approach intended to autonomously detect Egyptian hieroglyphs directly from images, such as those captured by robots. Additionally, our method can precisely locate these hieroglyphs within the image and provide a prediction percentage for each detected hieroglyphs.

Figure 1 depicts the proposed method.

To develop a proficient deep learning model for

Lombardi, L., Mercaldo, F. and Santone, A.
Egyptian Hieroglyphs Localisation Through Object Detection.
DOI: 10.5220/0012787900003753

In *Proceedings of the 19th International Conference on Software Technologies (ICSOFT 2024)*, pages 434-441
ISBN: 978-989-758-706-1; ISSN: 2184-2833

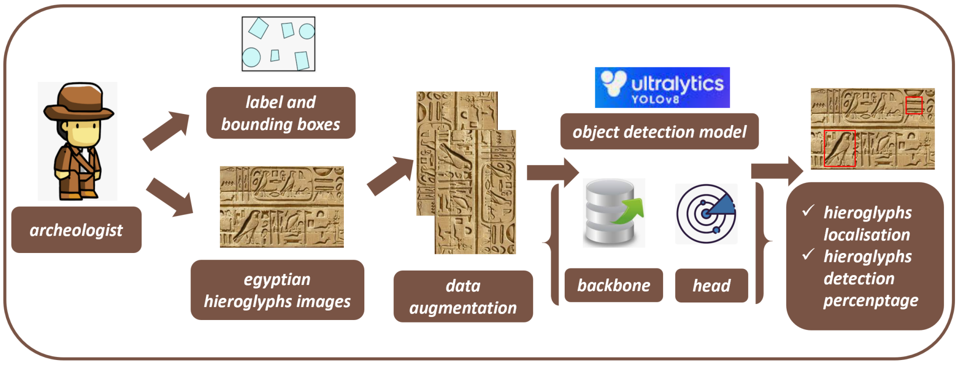

Figure 1: The proposed method.

detecting Egyptian hieroglyphs in images, it's crucial to possess a dataset containing images sourced from expert archeologists as well as robots or drones (refer to "archeologist" in Figure 1). Robots or drones are especially valuable for capturing images in remote or difficult-to-access locations. In order to construct a model capable of not only identifying Egyptian hieroglyphs within an image but also accurately locating them, a dataset containing images with detailed hieroglyph positions is indispensable. These images are meticulously labeled and annotated by domain experts (i.e., "archeologist" in Figure 1) to mark the areas in the images where the Egyptian hieroglyphs are present (referred to as "label and bounding boxes" in Figure 1). The detection class for the bounding box is singular, focusing solely on the ankh symbol (indicating the proposed model's emphasis on detecting fragments of the ankh Egyptian hieroglyph). The choice to prioritize the detection of this specific hieroglyph is due to its historical significance as a widely used decorative motif in ancient Egypt and neighboring cultures, symbolizing "life" itself.

Furthermore, to ensure the development of an effective model capable of accurately predicting unseen images, a diverse set of images captured from various angles and under different conditions is necessary. While these images may vary in size initially, a preprocessing step is required to standardize their dimensions.

The subsequent phase involves augmenting the dataset, as depicted in Figure 1 ("image augmentation"). This augmentation process utilizes a range of techniques to expand the dataset without the need for additional data collection. By introducing controlled random alterations to existing images, such as flips, augmented duplicates are generated. This technique enhances the precision of artificial neural networks during the training process by exposing them to a broader range of data.

Specifically, data augmentation techniques are employed to create images with controlled random changes, such as flips (see (Shorten and Khoshgoftaar, 2019)). This approach aims to ensure the model's effectiveness in recognizing Egyptian hieroglyphs regardless of their position within the image. Additionally, augmented data helps mitigate the risk of overfitting by preventing the model from becoming overly tailored to specific instances encountered during training.

Once the (augmented) images, along with associated details regarding the hieroglyph class and bounding boxes, are obtained, the next step is to implement a deep learning model (referred to as the "Object Detection model" in Figure 1).

In this paper, we adopt the YOLO 8 model ((Redmon et al., 2016)), which revolutionized object detection with its one-stage deep learning approach. YOLO is designed to classify images and accurately determine the positions of detected objects within them.

The key feature distinguishing YOLO from other models is its ability to perform the entire detection process in a single step. The YOLO process involves inputting an image and producing an output consisting of two components: a bounding box vector associated with the predicted class of the detected object for each cell in a grid representing the image.

Each image is divided into a grid of cells, and a cell is responsible for an object if the object's center falls within it. The bounding box prediction includes five components: (x, y, w, h, confidence), where (x, y) represent the box's center normalized to the cell's position, and (w, h) represent the box's dimensions normalized to the image size. Consequently, the predictions consist of S x S x B * 5 outputs for the bounding boxes ((Jiang et al., 2022)).

Compared to existing models, YOLO offers significantly faster performance ((Sanchez et al., 2020; Sah et al., 2017)), thanks to its single-phase approach, which predicts bounding boxes, object probabilities, and classes without multiple sequential steps.

We choose YOLO for its speed and lower likelihood of identifying false positives in the image background compared to alternative models ((Horak and Sablatnig, 2019; Jiang et al., 2022)). These qualities make YOLO one of the most effective convolutional neural network models for object detection.

The YOLO network consists of a backbone, responsible for collecting and organizing image features, and a Head, which utilizes these features for box and class prediction. Between the backbone and the head lies the neck, which integrates image features before forwarding them for prediction. In this paper, we experiment with the YOLOv8s model, the smaller version of YOLO 8 ((Yan et al., 2022)).

3 THE EXPERIMENT

In this section, we showcase the outcomes of our experimental investigation, which was conducted to demonstrate the efficacy of employing the YOLO 8 model for detecting and localizing Egyptian Hieroglyphs within images.

We compiled images from the *COTA_COCO_anks Image Dataset*, a collection specifically curated for constructing models geared towards Egyptian Hieroglyph detection from images. This dataset is openly accessible for research endeavors[1].

The utilized dataset comprises 1729 distinct real-world images featuring Egyptian Hieroglyphs. Each image is annotated with a single label, specifically "ank," along with its corresponding localization, represented by a bounding box indicating the position of the Egyptian Hieroglyph within the image. To ensure the reproducibility of our results, we opted to utilize a publicly available dataset.

The images of Egyptian Hieroglyphs are stored in JPEG format with a resolution of 640 x 640 pixels. We divided the images as follows: 1230 images for training, 325 for validation, and the remaining 174 for the test set, representing an approximate split percentage of 70:20:10, respectively. The dataset we acquired comes with annotations, meaning that each image includes detailed information about the bounding box surrounding each Egyptian Hieroglyph. Image augmentation was performed by generating additional images for each original image through horizontal and vertical flips. After implementing data augmentation, we obtained the final dataset. For model training, we selected a batch size of 16 and set the number of epochs to 10, with an initial learning rate of 0.01. For model training and testing, we utilized a machine equipped with an NVIDIA Tesla T4 GPU card boasting 16 GB of memory.

The experimental results obtained from our proposed method are visualized in Figures 2 through various plots.

Figures 2 present the experimental findings derived from the proposed method, showcased through multiple plots. In the top row of plots in Figure 2, the following metrics are depicted: "train/box_loss" (representing the trend of box_loss during training, measuring the fidelity of predicted bounding boxes to the ground truth object), "train/obj_loss" (illustrating the trend of obj_loss during training, where objectness determines the presence of an object at an anchor), "train/cls_loss" (displaying the trend of cls_loss during training, which gauges the accuracy of object classification within each predicted bounding box, with each box potentially containing an object class or "background" – commonly referred to as cross-entropy loss), precision trend, and recall trend. In the bottom row of plots in Figure 2, the following metrics are showcased: "val/box_loss" (depicting the trend of box_loss in validation), "val/obj_loss" (illustrating the trend of obj_loss in validation), mean Average Precision when Intersection over Union is equal to 0.5 (mAP_0.5), and mean Average Precision when Intersection over Union ranges between 0.5 and 0.95 (mAP_0.5:0.95).

All the metrics demonstrate the anticipated patterns: precision, recall, mAP_0.5, and mAP_0.5:0.95 are expected to increase as the number of epochs progresses, indicative of the model effectively learning to detect objects from Egyptian hieroglyphs images. Conversely, the other metrics display a decreasing trend as the number of epochs increases, providing further evidence that the model is effectively learning from Egyptian hieroglyph images. Specifically, the loss metrics generally signify instances where the model misidentifies a particular object, hence the loss values typically start high in the initial epochs and gradually decrease as the model learns to accurately detect objects of interest.

In Figures 2, the plots for metrics/mAP_0.5 and metrics/mAP_0.5:0.95 illustrate the mAP value for IOU=50 and IOU ranging from 50 to 95 (signifying different IoU thresholds from 0.5 to 0.95, with a step size of 0.05) on average mAP.

It is noteworthy that both the metrics/mAP_0.5 and metrics/mAP_0.5:0.95 plots in Figure 2 exhibit an increasing trend. This observation indicates that the model is effectively learning the spatial locations within images to accurately identify the objects of in-

[1]https://universe.roboflow.com/matthew-custer-bclqa/cota_coco_anks/dataset/3

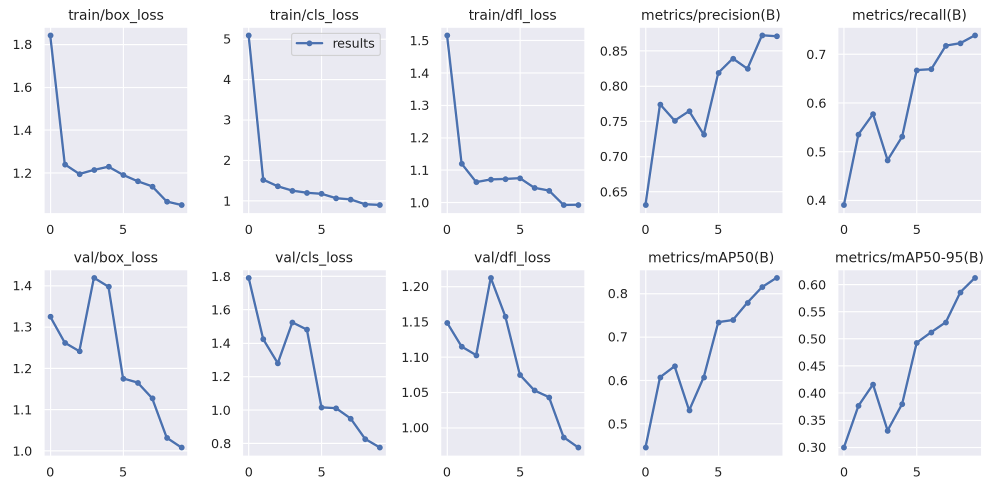

Figure 2: Experimental analysis results.

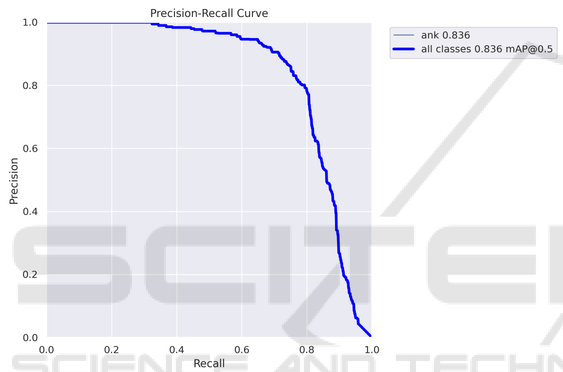

Figure 3: The Precision-Recall graph.

terest, namely humans and dogs.

Table 1 presents the results for Precision, Recall, mAP_0.5, and mAP_0.5:0.95 metrics, provided individually for the ank Egyptian hieroglyph class.

Table 1: Classification results.

Precision	**Recall**	**mAP_0.5**	**mAP_0.5:0.95**
0.87074	0.73876	0.83606	0.61249

Observing Table 1, we observe that the Precision and Recall values are 0.87 and 0.738, respectively for the ank Egyptian hieroglyph class.

Additionally, for a more comprehensive assessment of the proposed method, precision and recall values are visualized on the Precision-Recall graph in Figure 3.

The expected trend of this plot is a monotonically decreasing one, as there is typically a trade-off between precision and recall: increasing one usually results in a decrease in the other. While there can be exceptions or data limitations that prevent the precision-recall graph from strictly following this trend, the plot in Figure 3 demonstrates a decreasing trend for the relevant labels.

Moreover, the precision-recall plot displays the Area Under the Curve (AUC) values associated with the analysed class (i.e., ank Egyptian hieroglyph), along with the identification the ank class with mmAP_0.5. As previously mentioned, the precision-recall trend is anticipated to be monotonically decreasing. This behavior is evident from the precision-recall plot concerning all classes with mAP_0.5, yielding an AUC of 0.836. Given that these metrics range from 0 to 1, these values serve as indicative evidence that the proposed model is capable of effectively detecting ank Egyptian hieroglyphs from images.

To visually verify the proposed method and affirm its efficacy in real-world scenarios, we present examples of images with manually performed annotations (depicted in Figure 4) and the same images with detections and bounding boxes (illustrated in Figure 5). This enables a direct comparison between the manual annotations and those generated by the proposed model.

As depicted in Figure 4, we incorporate images captured from various angles and with subjects positioned at varying distances. This approach aims to maximize the model's generalizability. Notably, the proposed model demonstrates the capability to detect more ank Egyptian hieroglyphs within the same images, unaffected by background color. In Figure 4, each image includes detailed bounding box annotations for the classes involved in the experiment (highlighted in red) along with their respective label i.e., ank.

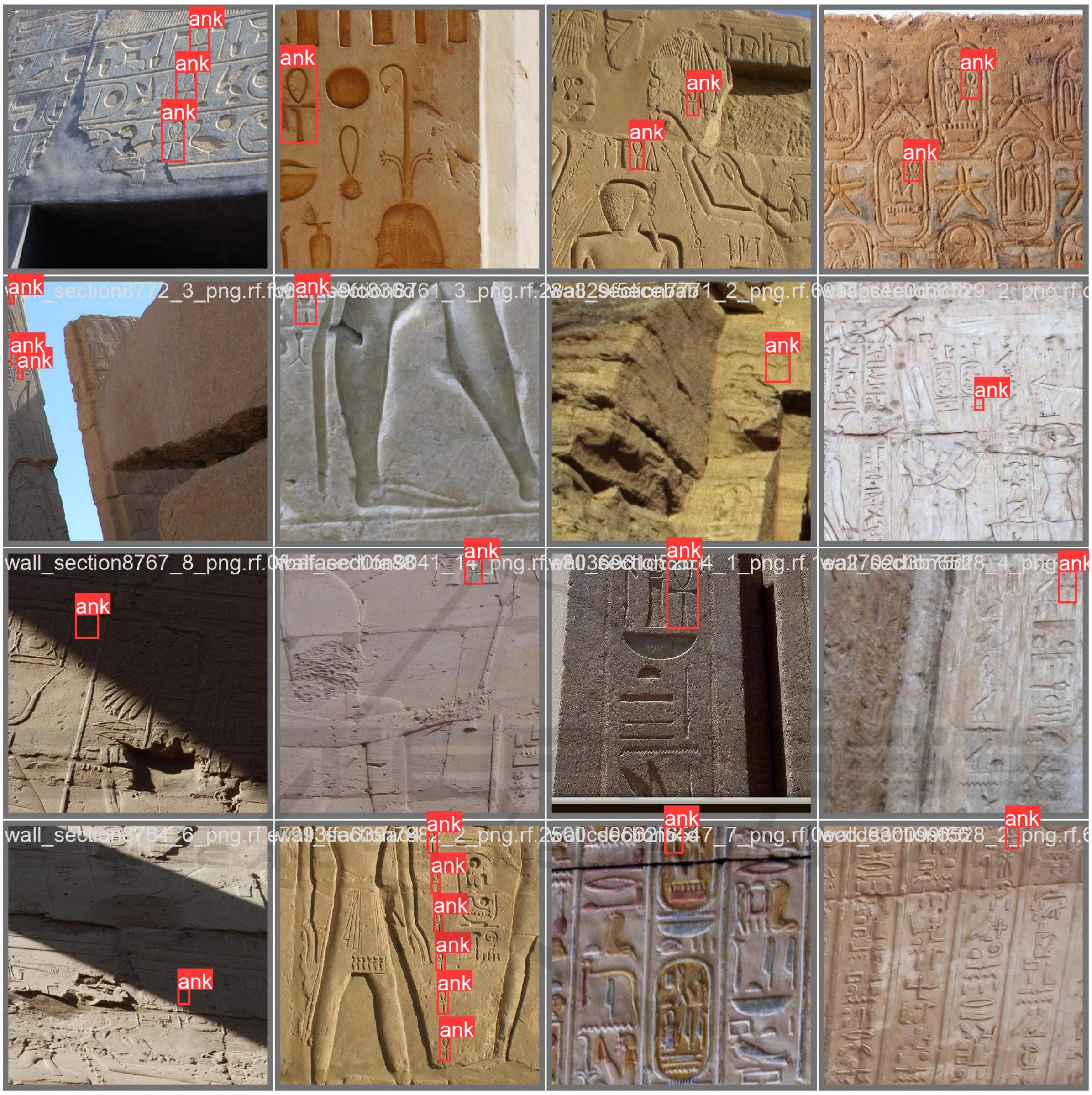

Figure 4: Example of images with the related bounding box around the ank Egyptian hieroglyphs, manually added for model building.

Figure 5 displays the predictions and bounding boxes generated by the proposed model during the testing phase. It is important to note that during testing, the images are inputted to the model without any pre-existing bounding boxes.

The observations from Figure 5 reveal that the proposed model demonstrates proficient localization of areas containing anks. In fact, for the majority of the images, the bounding boxes closely resemble those depicted in Figure 4.

As can be observed from the images displaying both ank manual annotations and ank annotations predicted by the developed model (i.e., Figure 4 and Figure 5), the background does not appear to be a disrupting factor for the model. In fact, the ankhs are correctly identified even when present alongside other hieroglyphs on materials of the same color. Moreover, even under low-light conditions and varying angles, the ankhs are accurately identified. The size of the ankh also does not pose an issue; as seen from the examples, regardless of size, the hieroglyph is correctly identified and notably not confused with other hieroglyphs typically present in the images.

Figure 5: Example of ank predictions automatically performed by the proposed method.

4 RELATED WORK

The problem of ancient Egyptian language retrieval and classification has been addressed, with different purposes, in several works and several examples of applications of the new technologies to the classification of ideograms belonging to ancientor no more used languages can be found in the literature. In this section, we review the current state-of-the-art literature about the application of deep learning models on the DR detection. Below we discuss these papers.

Authors in (Barucci et al., 2021), used ResNet-50 to develop a classification method to classify the glyphs. They didn't find the sufficient dataset to train their approach, so they used a different dataset then they used transfer learning. They also implemented a novel architecture called Glyphnet and trained it on a small hieroglyphic dataset which is designed for the specific task of hieroglyph classification and trained the network on it. The result showed that Glyphnet achieved an accuracy rate of 96% which is the highest accuracy found literature. But the data used in their work is unfortunately not available for other researchers to validate the results.

Researchers in (Moustafa et al., 2022), show Deep learning techniques, such as EfficientNet, MobileNet, and ShuffleNet. This study has been tested on two hieroglyph datasets. This paper describes a flutter-based mobile application named Scriba. This application provides as an advantage an exact translation of hieroglyphs.

Authors in (Hamdany et al., 2021),presents a novel method in 2021 for translating cuneiform writing by utilizing ANN. For all intents and purposes, a Multi-Layer Perceptron (MLP) neural network has been adapted to translate images of Sumerian cuneiform symbols into corresponding English letters. The process involves utilizing an artificial intelligence technique, the MLP neural network, to translate images of Sumerian cuneiform symbols into English letters. The central concept behind the method that authors have suggested is to acquire an image of any cuneiform symbol and then generate an indicator for the corresponding English letter. Considering this, it is possible for letters of the English alphabet to be intelligently generated from cuneiform symbol images. The proposed method uses neural networks, but it can also be represented using other information technology and artificial intelligence models. This work has been successfully established, and it achieved a score of %100; however, it faced challenges such as the inability to recognize chipped circuit boards, lack of public access to the dataset, difficulty in determining network structure, and absence of pre-processing. Consequently, the proposed network is unable to deal with the variations that can be found in images, such as rotation, noise, and so on.

In 2023, authors in (Mohsen et al., 2023), developed the idea of of Aegyptos: Mobile Application for Hieroglyphs Detection, Translation and Pronunciation, was brought up to help learn how to read Hieroglyphs and also help to pronounce them by using a tool at the palm of users' hands with just their phones' live cameras. Performing segmentation on the symbols using different segmentation techniques like Otsu Thresholding, while a lightweight CNN known as SqueezeNet is used for classification with the help of an API to translate the script into a language understood by the user.

5 CONCLUSION AND FUTURE WORK

In this work, we have explored the capability of deep learning techniques to face the problem of ancient Egyptian hieroglyphs classification and analysis of results was conducted to determine the champion models and the best data settings. Our approach allows DL to learn representations of images with better generalization performance, enabling the discovery of targets that have been difficult to find in the past. Moreover, by accelerating the research process, our method contributes to archaeology by establishing a new paradigm that combines field surveys and AI, leading to more efficient and effective investigations. Even though in this paper we focused on the single hieroglyph classification task, new and profitable perspectives are opened by the application of deep learning techniques in the Egyptologic field. In this view, the proposed work can be seen as the starting point for the implementation of much more complex goals. In future work, the incorporation of other hieroglyphs is proposed to expand the algorithm's database.The approach would be beneficial for the future of archaeology in a new paradigm of combining field survey and AI.

ACKNOWLEDGMENT

This work has been partially supported by EU DUCA, EU CyberSecPro, SYNAPSE, PTR 22-24 P2.01 (Cybersecurity) and SERICS (PE00000014) under the MUR National Recovery, Resilience Plan funded by the EU - NextGenerationEU projects, by MUR - REASONING: foRmal mEthods for computAtional analySis for diagnOsis and progNosis in imagING - PRIN, e-DAI (Digital ecosystem for integrated analysis of heterogeneous health data related to high-impact diseases: innovative model of care and research), Health Operational Plan, FSC 2014-2020, PRIN-MUR-Ministry of Health and the National Plan for NRRP Complementary Investments D^3 4 Health: Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care and Progetto MolisCTe, Ministero delle Imprese e del Made in Italy, Italy, CUP: D33B22000060001, and by Fondazione Intesa SanPaolo Onlus in the "Doctorates in Humanities Disciplines" for the "Artificial Intelligence for the Analysis of Archaeological Finds" topic.

REFERENCES

Barucci, A., Cucci, C., Franci, M., Loschiavo, M., and Argenti, F. (2021). A deep learning approach to ancient egyptian hieroglyphs classification. *Ieee Access*, 9:123438–123447.

Barucci, A. and Neri, E. (2020). Adversarial radiomics: the rising of potential risks in medical imaging from adversarial learning. *European Journal of Nuclear Medicine and Molecular Imaging*, 47(13):2941–2943.

Hamdany, A. H. S., Omar-Nima, R. R., and Albak, L. H. (2021). Translating cuneiform symbols using artificial neural network. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 19(2):438–443.

Horak, K. and Sablatnig, R. (2019). Deep learning concepts and datasets for image recognition: overview 2019. In *Eleventh international conference on digital image processing (ICDIP 2019)*, volume 11179, pages 484–491. SPIE.

Huang, P., Xiao, H., He, P., Li, C., Guo, X., Tian, S., Feng, P., Chen, H., Sun, Y., Mercaldo, F., et al. (2024). La-vit: A network with transformers constrained by learned-parameter-free attention for interpretable grading in a new laryngeal histopathology image dataset. *IEEE Journal of Biomedical and Health Informatics*.

Jiang, P., Ergu, D., Liu, F., Cai, Y., and Ma, B. (2022). A review of yolo algorithm developments. *Procedia Computer Science*, 199:1066–1073.

Mercaldo, F., Zhou, X., Huang, P., Martinelli, F., and Santone, A. (2022). Machine learning for uterine cervix screening. In *2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 71–74. IEEE.

Mohsen, S. E., Mansour, R., Bassem, A., Dessouky, B., Refaat, S., and Ghanim, T. M. (2023). Aegyptos: Mobile application for hieroglyphs detection, translation and pronunciation. In *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 1–8.

Moustafa, R., Hesham, F., Hussein, S., Amr, B., Refaat, S., Shorim, N., and Ghanim, T. M. (2022). Hieroglyphs language translator using deep learning techniques (scriba). In *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 125–132. IEEE.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

Sah, S., Shringi, A., Ptucha, R., Burry, A. M., and Loce, R. P. (2017). Video redaction: a survey and comparison of enabling technologies. *Journal of Electronic Imaging*, 26(5):051406.

Sanchez, S., Romero, H., and Morales, A. (2020). A review: Comparison of performance metrics of pretrained models for object detection using the tensorflow framework. In *IOP Conference Series: Materials Science and Engineering*, volume 844, page 012024. IOP Publishing.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.

Yan, T., Sun, W., and Cui, K. (2022). Real-time ship object detection with yolor. In *Proceedings of the 2022 5th International Conference on Signal Processing and Machine Learning*, pages 203–210.

Zhou, X., Tang, C., Huang, P., Tian, S., Mercaldo, F., and Santone, A. (2023). Asi-dbnet: an adaptive sparse interactive resnet-vision transformer dual-branch network for the grading of brain cancer histopathological images. *Interdisciplinary Sciences: Computational Life Sciences*, 15(1):15–31.