Article

Egyptian Hieroglyphs Segmentation with Convolutional Neural Networks

Tommaso Guidi, Lorenzo Python, Matteo Forasassi, Costanza Cucci, Massimiliano Franci, Fabrizio Argenti and Andrea Barucci

Topic Collection

Traditional and Machine Learning Methods to Solve Imaging Problems

Edited by
Dr. Laura Antonelli and Dr. Lucia Maddalena

Article

Egyptian Hieroglyphs Segmentation with Convolutional Neural Networks

Tommaso Guidi [1,2], **Lorenzo Python** [2], **Matteo Forasassi** [2], **Costanza Cucci** [1], **Massimiliano Franci** [3], **Fabrizio Argenti** [2] and **Andrea Barucci** [1,*]

1   Institute of Applied Physics "Nello Carrara", National Research Council, Via Madonna del Piano 10, Sesto Fiorentino, 50019 Firenze, Italy
2   Department of Information Engineering, University of Florence, Via S. Marta 3, 50139 Firenze, Italy
3   Center for Ancient Mediterranean and Near Eastern Studies, CAMNES, Via del Giglio 13, 50123 Firenze, Italy
*   Correspondence: a.barucci@ifac.cnr.it

Abstract: The objective of this work is to show the application of a Deep Learning algorithm able to operate the segmentation of ancient Egyptian hieroglyphs present in an image, with the ambition to be as versatile as possible despite the variability of the image source. The problem is quite complex, the main obstacles being the considerable amount of different classes of existing hieroglyphs, the differences related to the hand of the scribe as well as the great differences among the various supports, such as papyri, stone or wood, where they are written. Furthermore, as in all archaeological finds, damage to the supports are frequent, with the consequence that hieroglyphs can be partially corrupted. In order to face this challenging problem, we leverage on the well-known Detectron2 platform, developed by the Facebook AI Research Group, focusing on the Mask R-CNN architecture to perform segmentation of image instances. Likewise, for several machine learning studies, one of the hardest challenges is the creation of a suitable dataset. In this paper, we will describe a hieroglyph dataset that has been created for the purpose of segmentation, highlighting its pros and cons, and the impact of different hyperparameters on the final results. Tests on the segmentation of images taken from public databases will also be presented and discussed along with the limitations of our study.

Keywords: deep learning; convolutional neural networks; image recognition; ancient Egyptian hieroglyphs; cultural heritage

1. Introduction

The applications of the *Artificial Intelligence* (AI) are permeating almost any field of science, from chemistry and physics [1,2] to healthcare [3] and space research [4]. In recent years, this exponentially growing pattern can be primarily attributed to the particular class of algorithms known as Neural Networks or Deep Learning [5] (a sub-field of AI), which are adaptable to a variety of situations and well-suited to a data-driven approach to problem solving. More generally, the success of these methods has been also made possible by the advances in both computational power and telecommunications, which allow developers to share huge datasets and computational resources [6,7].

Computer vision is undoubtedly one of the scientific disciplines that has been most impacted by AI's advancements; currently, the vast majority of its applications are based on a specific class of Deep Neural Networks called *Convolutional Neural Networks* (CNN). Many of the issues in this subject can be solved by CNNs, including detection, recognition and segmentation tasks, as well as images generation [8–11].

The deployment of large neural network architectures capable of facing such tasks resulted in a horizontal spread of these technologies across many different fields, such as clinical imaging [12–14] or cultural heritage [15]. In the latter field, solutions were proposed, for example, for the recognition of the old Kuzushiji Japanese writing style [16,17], Maya [18] and Egyptian hieroglyphs classification [19,20]. These applications can be a very

challenging test for CNNs, due to the complexity of the problem deriving from the artifacts' state of conservation or the high variability of documents [21–23]. Some of these challenges will be discussed later in this article.

In this work, we evaluate the application of the well-known framework Detectron2 [24–26] when applied to the study of ancient Egyptian hieroglyphs, with the final goal of offering a solution to the segmentation task. Problems related to the creation of a well-suited dataset to train a CNN are described. In addition, some limitations, mainly related to the dimensions and quality of images, are also pointed out.

Segmentation is a fundamental step in several hieroglyphs analysis tasks: for example, it is the starting point for symbol transliteration and, subsequently, for language translation; at the same time, it may open the possibility to study the similarities between glyphs, moving towards the identification of the so-called "hand of the scribe". Therefore, another objective of this study may be seen as investigating the role of CNNs supporting the work of Egyptologists.

The paper is organised as follows: In Section 2, we address the complexity of the hieroglyphic writing system, in order to highlight the challenging aspects that any Deep Learning application might encounter. Then, in Section 3, we describe the main features and the reasons why we opted to use Detectron2, followed by an explanation of our dataset creation procedure. In Sections 4 and 5, we provide the results obtained on the validation and test sets and on images taken from public datasets. Concluding remarks are presented in Sections 6.

2. Ancient Egyptian Hieroglyphs

Ferdinand de Saussure defines the Egyptian word as a linguistic sign with a *signifier*, which is the graphic aspect, and a *signified*, the linguistic internal structure [27]. The *signifier* can be composed of one or many hieroglyphs at once. The Egyptian hieroglyph itself is composed of two different elements: the phonogram and the semagram. The former plays a phonetic role: it could either indicate the sound of the sign (or of a sequence of signs), or act as a phonetic complement. The latter is the graphic symbol, expressing an idea which is related to it. Inside a word, the semagram could simply act as the represented object indicating a word, or as a determinative, specifying the lexical field the word belongs to.

Ancient Egyptian hieroglyphs are represented by more than 700 ideograms belonging to about 26 categories; these ideograms are combined to create words and sounds, as briefly explained before. There are different ways of writing hieroglyphs, such as *cursive* or *monumental*; they can be written in different directions as well, as shown in Figure 1.
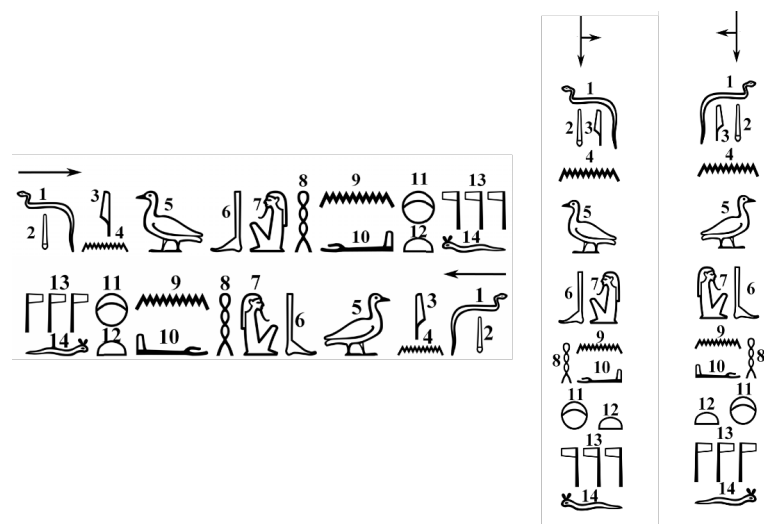

Figure 1. Various reading directions of ancient Egyptian hieroglyphs.

Another very important point to bear in mind is that hieroglyphs were written on a large variety of supports, such as papyri, stone and wood, with significant differences: while in the papyrus case, we deal with shapes filled of ink applied on the support, in the case of stone hieroglyphs being carved, thus forcing the segmentation algorithm to base its recognition power on the shadows, which can change depending on the lighting conditions and set-ups adopted during the image acquisition. Furthermore, when working with artifacts dating back thousands of years, the state of conservation may become critical, with many hieroglyphs potentially damaged and, consequently, a loss of information must always be taken into account, since ageing of the surfaces causes fading and alteration of the symbols

The highly complex scenario just depicted is a very intriguing test for CNNs, which are able to "learn" from data and incorporate into a network the knowledge of the "expert" without attempting to construct a model (which is too difficult to be estimated). At the same time, non-trivial problems must be faced when a Deep Learning solution is pursued: the principal issues are related to data scarcity and sparsity.

3. Materials and Methods

This work aims to provide a solution to the problem of the segmentation of all the instances of Egyptian hieroglyphs in a given image; in this section, the deep learning tools and the used dataset are described.

3.1. Detectron2

The problem of multiple instances segmentation in an image by using Deep Learning approaches has been extensively studied in recent years, as evidenced in several review papers [28,29]. For the specific task of image segmentation, many CNN architectures have been proposed, such as the popular U-Net [12], mainly devoted to semantic segmentation, Fast R-CNN and Faster R-CNN [10,11]. Mask R-CNN [24], an extension of the Faster R-CNN, is well suited for instances segmentation.

In this study, we used Detectron2 [25,26], a framework developed by the Facebook AI Research Group (FAIR). One of the main reasons for choosing Detectron2 is that it represents a "state of the art" framework for the task of image segmentation, achieved as an evolution of previous networks. Detectron2 includes implementations for many object detection algorithms, such as the above-mentioned Mask R-CNN, Faster R-CNN and Fast R-CNN. An advantage of Detectron2 is its flexibility in integrating different networks as the backbone to be used for subsequent processing steps, e.g., classification (which, however, were not considered in this study). We decided to use Mask R-CNN as object detection algorithm in the Detectron2 framework, relying on the results coming from benchmark studies, such as [29–32], which demonstrated the high performances of Mask R-CNN in comparison to other networks.

More specifically, region-based CNNs are a family of Convolutional Neural Networks that tries to extend classical detection and recognition tasks to the case of images with multiple instances of objects of interest [10,11]. In a nutshell, the system consists of two modules. The first module is the Region Proposal Network (RPN), which is responsible for generating category-independent *region* proposals, i.e., areas of the input image that are the candidate to contain an object of interest. These regions are also referred to as *Regions of Interest* (RoI). An *objectness score* is associated with each RoI: this score quantifies the probability that an area contains an object belonging to one of the classes of interest *vs.* background. The second module classifies the proposed regions. The two modules share some convolutional layers in order to optimize calculations and rely upon a single deep CNN (the *backbone*), which is responsible for generating the feature maps. In this study, we selected as a backbone network the well-known ResNet [8]; more specifically, we used a model that was pre-trained on the COCO dataset [33] and that was finely tuned over the dataset of segmented hieroglypghs (described in the following), exploiting the *transfer learning* paradigm. The output of the segmentation network is a set of rectangular

bounding boxes, hopefully one for each instance of an object belonging to one of the classes of interest, along with a label for the predicted class. Parallel to the classification task, Mask R-CNN also produces for each RoI, in a parallel branch, a *binary mask* that, in the ideal case, should be perfectly superimposed to the object contained in the bounding box. This binary mask is the actual outcome of the segmentation task we are interested in: it is worth noticing that this is performed independently from both bounding-box regression and classification [24].

3.2. The Dataset

Training the Mask R-CNN needs the availability of a suitable dataset containing a number of different Egyptian hieroglyphs (as large as possible), along with a segmentation mask associated with each of them (the *ground truth*). An amount of pictures are available online, for example in the repositories of museums, such as the collection of "Egyptian Museum in Turin" (Italy) [34] or "The Met" in New York (US) [35], but, unfortunately, all these images are not segmented. Furthermore, most of the images are not yet ready to be used for training, having different dimensions, resolutions, point of views, etc. Thus, likewise in many machine learning problems, the construction of a dataset for training purposes needs an accurate selection and preparation of the images. More specifically, better results are expected if the images contain symbols as uniform as possible (in terms of dimension). Often, images contain both hieroglyphs and drawings, so that, even though Detectron2 works with input images of arbitrary size, cropping patches from larger images was performed to extract "text" areas.

In this work, we used different resources to create our dataset. Part of the images comes from the dataset used by the authors of "GlyphReader" [36,37]. Images from this dataset represent the stelae contained in the Egyptian Pyramid of Unas, and each image contains many hieroglyphs. The number of images is limited (ten) and, more importantly for their use in a training process, they are quite similar to each other, i.e., they come from the same stelae and have the same support, so that they lack the "hetereogeneity" necessary to make the network learn the generic features important for segmentation. Therefore, the dataset was extended by adding images representing different writing styles and supports, available thanks to the online resources cited above. Our final dataset is then composed by 101 images, each containing multiple glyphs. The dataset was split in training (59 images), validation (14 images) and test set (28 images). Some examples are shown in Figure 2.

Figure 2. Examples of images (on different supports) belonging to our dataset: papyrus on the top and bas-relief on the bottom.

After image selection, all the instances in the training and validation sets were manually segmented and labeled using the *VGG Image Annotator* tool (VIA) [38] in order to create the "true" segmentation mask of all the hieroglyphs contained in the images. VIA allows images to be annotated with bounding-boxes, circles or, as in our case, polygonal masks. The annotations were saved in a *json* file, which can be read by the Detectron2 framework. At the end of the manual segmentation process, the gathered instances of hieroglyphs and associated masks for the training set were 2198 and those for the validation set were 564. An example of the outcome of this manual segmentation process is shown in Figure 3.
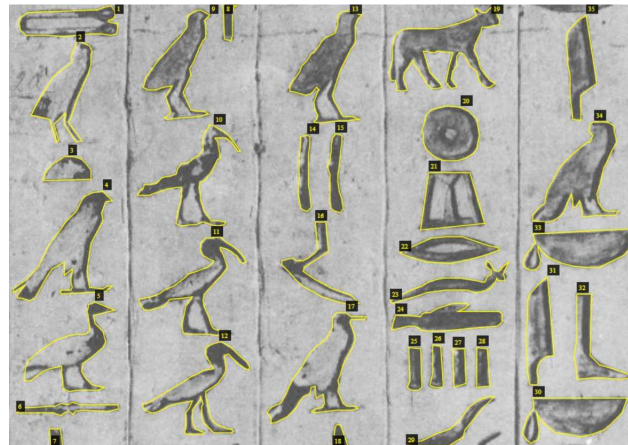

Figure 3. Examples of manual segmentation of the instances of hieroglyphs in an image using a VGG-Image Annotator.

Since in this work we were only interested in the segmentation task, we labeled each hieroglyph in an image as belonging to the generic *"glyph"* class, so that, at a pixel level, we are dealing with a binary classification task between *background* and *glyph*.

It is worth noting that hieroglyphs are represented by more than 700 different ideograms. The segmentation process benefits from using a training dataset populated with the majority of hieroglyph classes. In our case, it can be noticed that the classes of glyphs are not exhaustively represented and, in some cases, just a few instances per class are present. This fact can be an evident limitation to the generalization capability of our model. Despite that, in the Experimental Results section, we will show that the network is able to also segment glyphs that were not present into the training dataset.

4. Results on the Training and Validation Set

Several experiments were set up in order to evaluate the performance of the analysed network and its dependence on network hyper-parameters, such as thresholds and number of training epochs.

The network is trained to find the bounding-boxes and the polygonal of all the hieroglyphs within an image, without any distinction among the different classes of hieroglyphs. As already mentioned, this can be achieved by training the network to detect only one generic class, labelled as "glyph", from the surrounding "background". As an output of the segmentation process, we have an RoI with a mask highlighting the object and a bounding-box around it.

Two major choices influence the performance of the segmentation network: the threshold on the objectness score and the number of training epochs. The former is used by Mask R-CNN in order to select candidate regions to be fed to the backbone network for classification; thus, this threshold has an influence on the number of selected regions that can be classified as glyph. The latter influences the training duration and has an influence on the accuracy of the segmented regions; how to measure such an accuracy will be discussed in the following.

The effect of different choices of the hyperparameters is illustrated in Figure 4. Figure 4a,b show the output masks selected for two different values of the objectness score threshold when the number of epochs used during the training stage is 300. It is apparent that decreasing the threshold produces an increase of the number of selected objects, even though low quality instances are included (the higher the quality of an output mask, the closer its approximation to the underlying glyph). It is also apparent that high quality masks will be necessary when a classification (and then transliteration/translation) task will be associated with the segmentation one. The quality of the output masks can be enhanced by tuning the number of epochs of the training stage. Figure 4c, for example, shows the output masks extracted with a threshold equal to 0.5, but using 500 epochs.
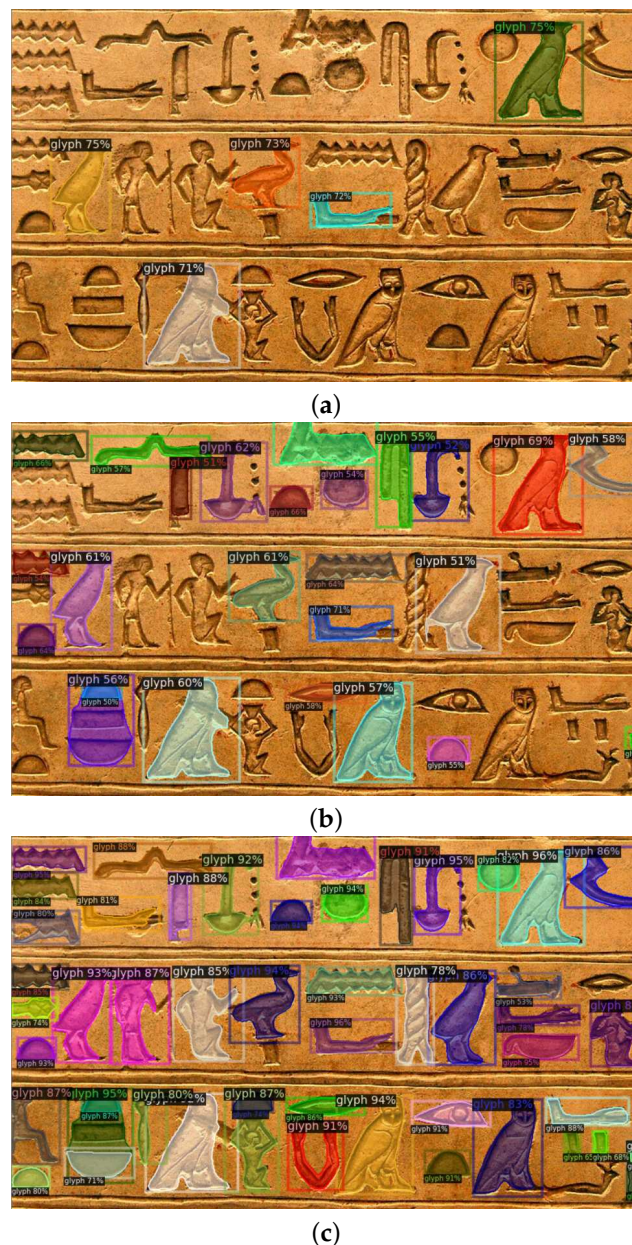

(**a**)

(**b**)

(**c**)

Figure 4. Segmentation results obtained after 300 epochs and objectness score threshold equal to 0.7 (**a**); after 300 epochs and objectness score threshold equal to 0.5 (**b**); and after 500 epochs and objectness score threshold equal to 0.5 (**c**).

A visual proof about the learning evolution of the network can be inferred from Figure 5, where we focus on the masks extracted for a single hieroglyph, with a threshold equal to 0.5, and after 300, 500 and 3000 epochs.
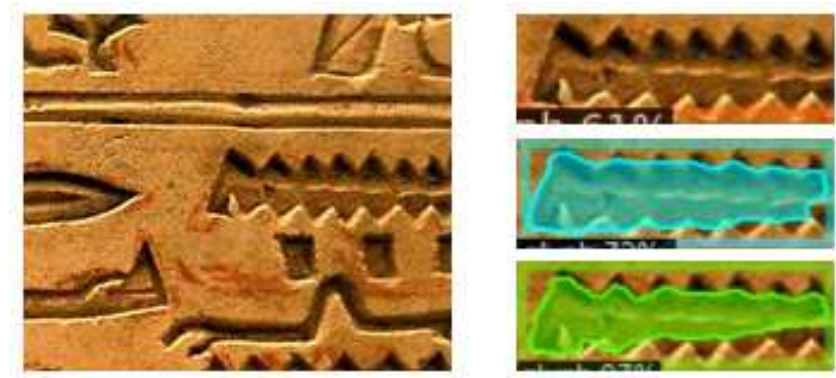

Figure 5. Focus on the segmentation of a single glyph (left) after training for 300, 500, 3000 epochs (right panel, from top to bottom).

The influence of network hyper-parameters is further illustrated in Figure 6, in which segmentation results for a large number of epochs (3000) and different objectness score thresholds are illustrated: Figure 6a,b refer to thresholds equal to 0.7 and equal to 0.1, respectively. From Figure 4a, obtained with a selective threshold (equal to 0.7) and a low number of epochs (equal to 300), we observe the detection of a low number of instances. Figure 6a demonstrates that increasing the number of epochs improves the objectness score so that the number of detected of instances considerably increases. Figure 6b, on the other hand, shows that a comparable quality, in terms of segmentation masks, can be achieved even with a poorly selective threshold (0.1).
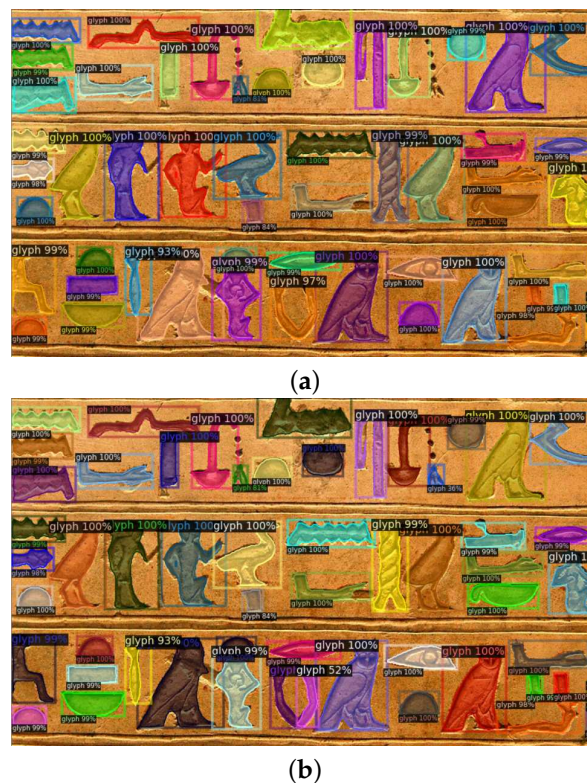

(**a**)

(**b**)

Figure 6. Segmentation results obtained with a training of 3000 epochs with an objectness score threshold equal to 0.7 (**a**) and with an objectness score threshold equal to 0.1 (**b**).

All the results shown so far were obtained starting the learning process from a model pre-trained on the COCO-dataset. To illustrate the importance of the transfer learning paradigm, we also tried to train the network "from scratch", achieving, however, significantly worse results. The segmentation obtained by using training from scratch with a duration of 3000 epochs and a threshold on the objectness score equal to 0.7 is shown in Figure 7. Despite the high confidence of the RPN, related to the high objectness scores, the quality of the segmentation masks is rather poor with respect to the results obtained with pre-training (see Figure 6a).
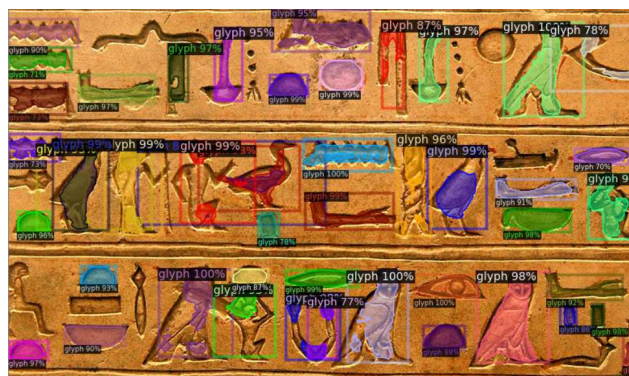

Figure 7. Segmentation results obtained with a training from scratch of 3000 epochs and with an objectness score threshold equal to 0.7.

For a quantitative assessment of the segmentation process, we can use metrics commonly used for this purpose. Given a glyph with a known segmentation mask, i.e., the ground truth, and the mask predicted by the network, the *Intersection Over Union* (IoU) is the ratio between the overlapping areas of ground truth and prediction over the area of their union; this metric ranges from 0 to 1. By setting a threshold on the IoU, we can discriminate successful segmentations (IoU above the threshold) from unsuccessful ones. Taking the average successful rates over the whole dataset yields an average precision (AP) of the segmentation process and, therefore, a measure of its quality. The AP metric obviously depends on the threshold that we use to select the successful segmentation: for example, by using a threshold of 0.75, we obtain a metric referred to as AP75 (expressed as percentage). The IoU and AP metrics can be applied not only to masks, but also to bounding boxes.

In Table 1, the segmentation quality evaluated on the validation set obtained for an objectness score threshold equal to 0.5 and for a different duration of the training process is shown; more specifically, the AP50 and AP75 values are tabulated. Results in Table 1 show that both bounding box and mask regression achieve better performances for a higher number of epochs, that is, they benefit from a longer training.

Table 1. Performance metrics of the segmentation process (on the validation set) vs. the training epochs (objectness score threshold = 0.5).

Number of Epochs	Bounding-Box Regression		Mask Regression	
	AP50	AP75	AP50	AP75
300	82.0%	56.0%	77.8%	37.8%
500	90.4%	77.3%	86.8%	51.3%
3000	93.4%	85.0%	91.2%	65.8%

Additionally, in order to further quantify the IoU behaviour varying with the choice of the different hyper-parameters, we report in Table 2 the results relative to the single glyph representing the "owl" in the bottom left part of the image in Figure 4. As can be seen, the

threshold value, which impacts the detection ability, has not influence on the IoU, while the number of iterations improves the quality of the segmentation as measured by the IoU.

Table 2. Performance metrics of the segmentation process for a specific hieroglyph ("owl"), vs. the training epochs and objectness score threshold ([f.s.] refers to the training from a scratch case).

Number of epochs	300	300	500	3000	3000	3000 [f.s.]
Threshold	0.5	0.7	0.5	0.1	0.7	0.7
IoU	0.84	0.84	0.87	0.90	0.90	0.33

In summary, the above results show the influence of some important hyper-parameters of the network on the segmentation performance. As to the objectness score, this influences the symbol detection ability. On the other hand, the number of epochs influences the IoU metrics, so that training duration can be traded for segmentation quality.

5. Results on the Test Set

Once the network has been trained and validated, it is ready for its use as a predictor for the segmentation of images containing hieroglyphs, indicated as *test set*. In the test set, we inserted images coming from different resources and characterized by various supports, such as papyrus, stone and pottery. The idea is testing the CNN in a real scenario, despite this possibly being distant from the training conditions.

Figure 8 shows some results that can be considered as examples of good quality hieroglyph segmentation; the pictures come from the online Egyptian collection of the MET of New York [35]. The instances were selected by using a threshold equal to 0.7 and a training duration of 3000 epochs. In this case, the support is stone, with hieroglyphs in bas-relief; for these signs, the extracted masks represent the underlying glyphs with a good quality (since the the ground truth is not available, only a qualitative evaluation is attempted here). We note also that part of the glyphs were excluded due to an objectness score below the threshold.

Figure 8. Examples of segmentation on stones. Cat. numbers DP253181, EG548). Credits to the MET collection [35].

Another segmentation example is shown in Figure 9, where the Cyperus papyrus is depicted. In addition, in this case, the quality of the final segmentation can be considered good for most of the extracted hieroglyphs, even though it is also possible to appreciate the difficulties encountered with those instances of glyphs characterized by a low contrast.

Figure 9. Segmentation example on Cyperus papyrus. From the *Book of the Dead of Nebhepet, scribe of the necropolis* (cat. number 1768). Credit to the Egyptian Museum of Turin online collection [34].

Many other tests were performed on images depicting signs on different supports (papyrus, pottery, wood), showing similar results as those described above. It is interesting to notice the ability of the Mask R-CNN to catch also glyphs that were not present in the train set. This can be ascribed to the ability of the deeper layers of the CNN to learn basic features (e.g., contours) and associate them to the recognition of a generic "glyph" object.

Deep Neural Networks are extremely high-performance algorithms to discover patterns in data, but some limitations in their application can be pointed out. A first issue is that training should be performed on a considerable amount of samples, especially if the aim is creating a model with generalization capabilities. In this work, the dataset used to train the CNN model contained a moderately large, but not huge, number of images and manually segmented glyphs. Thus, the main limitation to the obtained results could be related to an inadequate number of samples.

Moreover, hieroglyphs can be written on different supports. For example, in the case of papyrus, the shape of the hieroglyph is fully determined by the ink painted on the support, whereas, in the case of bas-reliefs, shadows strongly contribute to the shape of the glyph, so that it may depend on the different environmental conditions (illumination, angle, etc.) in which the pictures were taken.

Finally, as in all archaeological objects and historical artefacts, damage to the support is frequent, with the consequence that hieroglyphs can be partially corrupted. These effects are, for example, visible in Figure 10, showing the segmentation output of a papyrus. Glyphs with a good contrast (i.e., where the ink has been preserved) seem to also produce a good segmentation result (see, for example, the blue mask on the top left in Figure 10), whereas glyphs with low contrast or evident scrapings are badly or partially segmented; in some cases, the segmentation process completely fails: different instances are wrongly connected or even artifacts are classified as glyphs (see the red mask in Figure 10).

Figure 10. Example of critical segmentation results (on Cyperus papyrus, from the Egyptian Museum of Turin online collection, *Book of the Dead of Nebhepet, scribe of the necropolis*, cat. number 1768): on the left the original document, on the right the segmented one.

6. Conclusions

In this work, we analyze the properties of the Detectron2 framework as a tool to face the problem of ancient Egyptian hieroglyphs segmentation. A pre-trained model was employed and finely tuned by using an ad-hoc dataset constructed on images taken from different sources. Hieroglyph segmentation is a difficult task due also to the variety of the signs and the possible corruption of symbols caused by millennial ageing. Notwithstanding these adversities, in our experimental tests, we have shown that a CNN-based approach may be a good candidate to solve the segmentation problem, with promising quantitative results on the validation set and qualitatively fine results on images taken "from the wild". Segmentation performance is limited by several issues inherent to the quality and amount of the raw available data, for which possible countermeasures—such as using data augmentation or image enhancement techniques—still need to be investigated.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CNN	Convolutional Neural Network
R-CNN	Region based Convolutional Neural Network
RPN	Region Proposal Network
Mask R-CNN	Mask Region based Convolutional Neural Network
COCO	Common Objects in Context

References

1. Mater, A.C.; Coote, M.L. Deep learning in chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545–2559. [CrossRef] [PubMed]
2. Barucci, A.; D'Andrea, C.; Farnesi, E.; Banchelli, M.; Amicucci, C.; de Angelis, M.; Hwang, B.; Matteini, P. Label-free SERS detection of proteins based on machine learning classification of chemo-structural determinants. *Analyst* **2021**, *146*, 674–682. [CrossRef] [PubMed]
3. Yu, K.; Beam, A.; Kohane, I. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2018**, *2*, 719–731. [CrossRef] [PubMed]
4. Kothari, V.; Liberis, E.; Lane, N.D. The Final Frontier: Deep Learning in Space. In Proceedings of the HotMobile '20: 21st International Workshop on Mobile Computing Systems and Applications, Austin, TX, USA, 3 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 45–49. [CrossRef]
5. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 14 December 2022).
6. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
7. Carleo, G.; Cirac, I.; Cranmer, K.; Daudet, L.; Schuld, M.; Tishby, N.; Vogt-Maranto, L.; Zdeborová, L. Machine learning and the physical sciences. *Rev. Mod. Phys.* **2019**, *91*, 045002. [CrossRef]
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
9. Girshick, R.B.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv* **2013**, arXiv:1311.2524.
10. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef] [PubMed]
12. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
13. Avanzo, M.; Porzio, M.; Lorenzon, L.; Milan, L.; Sghedoni, R.; Russo, G.; Massafra, R.; Fanizzi, A.; Barucci, A.; Ardu, V.; et al. Artificial intelligence applications in medical imaging: A review of the medical physics research in Italy. *Phys. Med.* **2021**, *83*, 221–241. [CrossRef]
14. Scapicchio, C.; Gabelloni, M.; Barucci, A.; Cioni, D.; Saba, L.; Neri, E. A deep look into radiomics. *Radiol. Med.* **2021**, *126*, 1296–1311. [CrossRef]
15. Cucci, C.; Barucci, A.; Stefani, L.; Picollo, M.; Jiménez-Garnica, R.; Fuster-Lopez, L. Reflectance hyperspectral data processing on a set of Picasso paintings: Which algorithm provides what? A comparative analysis of multivariate, statistical and artificial intelligence methods. In Proceedings of the Optics for Arts, Architecture, and Archaeology VIII. International Society for Optics and Photonics, Online, 8 July 2021; Volume 11784, p. 1178404.
16. Lamb, A.; Clanuwat, T.; Kitamoto, A. KuroNet: Regularized Residual U-Nets for End-to-End Kuzushiji Character Recognition. *SN Comput. Sci.* **2020**, *1*, 177. [CrossRef]
17. Clanuwat, T.; Lamb, A.; Kitamoto, A. KuroNet: Pre-Modern Japanese Kuzushiji Character Recognition with Deep Learning. In Proceedings of the 15th International Conference on Document Analysis and Recognition (ICDAR), Sydney, Australia, 20–25 September 2019; pp. 607–614.
18. Roman-Rangel, E.; Marchand-Maillet, S. Indexing Mayan hieroglyphs with neural codes. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 253–258. [CrossRef]
19. Barucci, A.; Cucci, C.; Franci, M.; Loschiavo, M.; Argenti, F. A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification. *IEEE Access* **2021**, *9*, 123438–123447. [CrossRef]
20. Barucci, A.; Canfailla, C.; Cucci, C.; Forasassi, M.; Franci, M.; Guarducci, G.; Guidi, T.; Loschiavo, M.; Picollo, M.; Pini, R.; et al. Ancient Egyptian Hieroglyphs Segmentation and Classification with Convolutional Neural Networks. In Proceedings of the International Conference Florence Heri-Tech: The Future of Heritage Science and Technologies, Florence, Italy, 16–18 May 2022; pp. 126–139.
21. Michelin, A.; Pottier, F.; Andraud, C. 2D macro-XRF to reveal redacted sections of French queen Marie-Antoinette secret correspondence with Swedish count Axel von Fersen. *Sci. Adv.* **2021**, *7*, eabg4266. [CrossRef] [PubMed]

22. Bickler, S.H. Machine Learning Arrives in Archaeology. *Adv. Archaeol. Pract.* **2021**, *9*, 186–191. [CrossRef]
23. Mantovan, L.; Nanni, L. The computerization of archaeology: Survey on artificial intelligence techniques. *SN Comput. Sci.* **2020**, *1*, 1–32. [CrossRef]
24. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
25. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2. 2019. Available online: https://github.com/facebookresearch/detectron2 (accessed on 14 December 2022).
26. Detectron2 Documentation. Available online: https://detectron2.readthedocs.io/en/latest/index.html (accessed on 14 December 2022).
27. de Saussure, F. *Course in General Linguistics*; Columbia University Press: New York, NY, USA, 2011.
28. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef]
29. Bharati, P.; Pramanik, A. Deep learning techniques—R-CNN to mask R-CNN: A survey. In *Computational Intelligence in Pattern Recognition*; Springer: Singapore, 2020; pp. 657–668.
30. Quoc, T.T.P.; Linh, T.T.; Minh, T.N.T. Comparing U-Net convolutional network with mask R-CNN in agricultural area segmentation on satellite images. In Proceedings of the 2020 7th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, Vietnam, 26–27 November 2020; pp. 124–129.
31. Durkee, M.S.; Abraham, R.; Ai, J.; Fuhrman, J.D.; Clark, M.R.; Giger, M.L. Comparing Mask R-CNN and U-Net architectures for robust automatic segmentation of immune cells in immunofluorescence images of Lupus Nephritis biopsies. In Proceedings of the Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XIX, Online, 6–11 March 2021; Volume 11647, pp. 109–115.
32. Alfaro, E.; Fonseca, X.B.; Albornoz, E.M.; Martínez, C.E.; Ramrez, S.C. A Brief Analysis of U-Net and Mask R-CNN for Skin Lesion Segmentation. In Proceedings of the 2019 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), Budapest, Hungary, 3–5 July 2019; pp. 000123–000126. [CrossRef]
33. Available online: https://cocodataset.org/#home (accessed on 14 December 2022).
34. Museo Egizio Collection. Available online: https://museoegizio.it/en/discover/collection/ (accessed on 14 December 2022).
35. MET Collection. Available online: https://www.metmuseum.org/about-the-met/collection-areas/egyptian-art (accessed on 14 December 2022).
36. Franken, M.; van Gemert, J. Automatic Egyptian Hieroglyph Recognition by Retrieving Images as Texts. In Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, Spain, 21–25 October 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 765–768. [CrossRef]
37. Franken, M. GlyphReader. GitHub. 2017. Available online: https://github.com/morrisfranken/glyphreader (accessed on 14 December 2022).
38. Dutta, A.; Gupta, A.; Zissermann, A. VGG Image Annotator (VIA). 2016. Available online: http://www.robots.ox.ac.uk/~vgg/software/via/ (accessed on 14 December 2022).