IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 22, 2021, accepted August 28, 2021, date of publication September 3, 2021, date of current version September 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3110082

A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification

ANDREA BARUCCI[1], COSTANZA CUCCI[1], MASSIMILIANO FRANCI[2], MARCO LOSCHIAVO[3], AND FABRIZIO ARGENTI[3], (Senior Member, IEEE)

[1]Institute of Applied Physics "Nello Carrara" (IFAC), Italian National Research Council (CNR), Sesto Fiorentino, 50019 Florence, Italy
[2]CAMNES, 50123 Florence, Italy
[3]Department of Information Engineering, University of Florence, 50139 Florence, Italy

Corresponding author: Fabrizio Argenti (fabrizio.argenti@unifi.it)

ABSTRACT Nowadays, advances in Artificial Intelligence (AI), especially in machine and deep learning, present new opportunities to build tools that support the work of specialists in areas apparently far from the information technology field. One example of such areas is that of ancient Egyptian hieroglyphic writing. In this study, we explore the ability of different convolutional neural networks (CNNs) to classify pictures of ancient Egyptian hieroglyphs coming from two different datasets of images. Three well-known CNN architectures (ResNet-50, Inception-v3 and Xception) were taken into consideration and trained on the available images. The paradigm of transfer learning was tested as well. In addition, modifying the architecture of one of the previous networks, we developed a specifically dedicated CNN, named Glyphnet, tailoring its complexity to our classification task. Performance comparison tests were carried out and Glyphnet showed the best performances with respect to the other CNNs. In conclusion, this work shows how the ancient Egyptian hieroglyphs identification task can be supported by the deep learning paradigm, laying the foundation for information tools supporting automatic documents recognition, classification and, most importantly, the language translation task.

INDEX TERMS Deep learning, convolutional neural networks, image recognition and classification, ancient Egyptian hieroglyphs, cultural heritage.

I. INTRODUCTION

Artificial Intelligence (AI) and machine learning applications are spreading in any field of science, from fundamental physics to natural language processing and clinical medicine, with everyday awesome results strongly impacting our life [1]–[5]. Despite the concerns related to its use [6]–[8], 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.

Taking advantage of the results coming from other fields, where methods related to AI and machine learning have already laid strong and deep roots, in this work the problem of ancient Egyptian hieroglyphs classification is addressed.

Several examples of applications of the new technologies to the classification of ideograms belonging to ancient or no more used languages can be found in the literature. In [9], [10], for example, the KuroNet network, based on a Unet architecture, is proposed to recognize the old Kuzushiji Japanese writing style. In [11], linear autoencoders are used to characterize local regions of complex shapes and are applied to indexing a collection of hieroglyphs from the ancient Maya civilization.

The advantages of such techniques are numerous also for the Egyptian philology, both at the synchronic and diachronic level [12]: the graphemic and hieroglyphics palaeography evolutions, the recognition of variants, the calculation of the logographic, syllabic and alphabetic percentage of hieroglyphic writing system, to name a few. The problem of ancient Egyptian language retrieval and classification has been addressed, with different purposes, in several works. In [13], image descriptors and image matching techniques are proposed to classify a database of more than 4000 hieroglyphs [14]. In [15], computer vision methods are used to identify hieroglyphs in fragments of Egyptian cartouches with the aim of contributing to a navigation system for

The associate editor coordinating the review of this manuscript and approving it for publication was Junxiu Liu.

museums. Computer vision methods for hieroglyph recognition are used also in [16]. A text information retrieval system, designed to work with Egyptian hieroglyphic texts, has been proposed in [17]. Further, the identification and transliteration of the signs is proposed in [18] and [19]. Recently, the world of hieroglyphs recognition has witnessed new interest from the Google team [20], [21]. However, to the best of authors' knowledge, there is still a lack of tools enabling a semi-automated approach to deciphering ancient Egyptian texts.

Hieroglyphs are represented by a wide spectrum of ideograms, generally assignable to about 26 categories, which are combined to give words and sounds [22]. They were written in different ways, such as monumental or cursive, in different directions (see Fig. 1), and on various supports such as papyrus, wood and stones. Today, a civilization lasting almost 5000 years has left a large amount of documents, which still need to be acquired, translated and interpreted. Here is exactly where AI comes into play, supporting ideograms classification and subsequently translation.

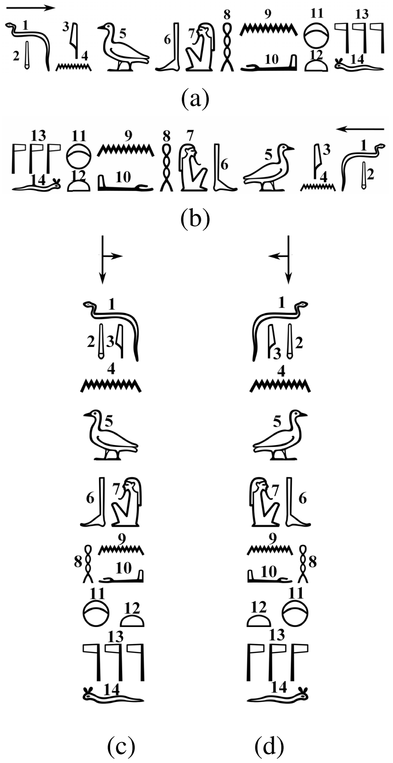


FIGURE 1. Examples of hieroglyphic signs and their reading direction.

In this work, we focused on single hieroglyph classification using the deep learning approach, in particular the architectures known as Convolutional Neural Networks (CNN), which can be considered the best choice for visual recognition tasks. Starting from two labelled datasets of ancient Egyptian hieroglyphs, one publicly available and the other constructed by the authors, three well-known CNNs – successfully proposed for image recognition tasks – were tested, either by using the transfer learning paradigm or by training from scratch. Inspired by the architecture of one of the previous networks, a new CNN, specifically tailored to the complexity of the classification problem at hand, was also developed. Experimental classification tests were performed to compare the classical CNNs and the new proposed one, referred to in the following as Glyphnet. Results demonstrate how deep learning methods yield extremely good results in terms of classification rates, with Glyphnet outperforming the other tested CNNs.

The paper is organized as follows. In Section II, some details about the ancient Egyptian hieroglyphic writing system and the used datasets are presented. In Section III, the CNNs that were tested to solve the classification task are described. The experimental tests that were performed to compare the various networks are shown in Section IV. Some concluding remarks are drawn in Section V.

II. THE PROBLEM OF HIEROGLYPHS RECOGNITION AND THE USED DATASETS

In this Section, the problem of hieroglyphs recognition is stated and the materials used in this study are described.

A. THE ANCIENT EGYPTIAN HIEROGLYPHIC WRITING SYSTEM

The Egyptian word is a linguistic sign with a signifier and a signified following the Sausserian definition [12]. The first component represents the external aspect, merely graphic, that can be composed of one or more hieroglyphics. The second one represents the internal structure, essentially linguistics. The Egyptian hieroglyph is a complex sign composed of two elements: a semagram (or ideogram) and a phonogram. The semagram is a graphic symbol representing an idea in relation to it. Our modern culture is literally surrounded by semagrams, just think of the road signs, the logos of many brands, or the social networks emoticons. A semagram can have two different values, depending on its function in the word: the proper semagram, which means the represented object indicating directly a word, and the determinative, a sign with a purely semantic and no phonetic value, whose function is to express the lexical field to which the word belongs. The "phonogram" may also have two different roles as well: the proper phonogram, which can indicate the phonetic value of the sign and metaphonically the sound (or phonetic sequence) only; and the phonetic complement, a specific series of signs that expresses in a redundant way the sound of the sign to which they are accompanied (see Fig. 2). Given this complex nature, the hieroglyphic sign proves to be a fertile ground for the application of a deep learning approach for its recognition and classification.

B. IMAGE DATASETS

In this study, two different datasets of images have been used, referred to in the following as D_1 and D_2 for the sake of brevity: the first one is publicly available [13] and

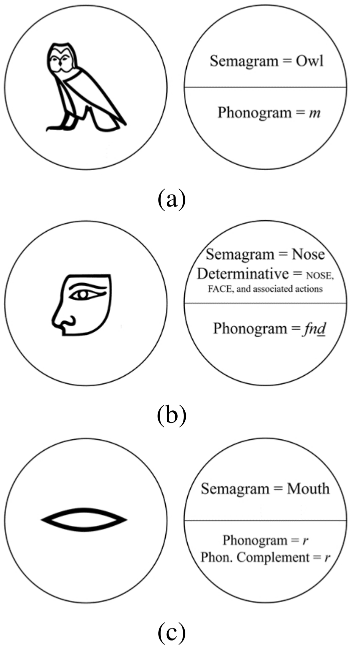


FIGURE 2. The structural elements of an Egyptian hieroglyph: (a) semagram (OWL) and phonogram /m/; (b) semagram (NOSE), determinative (nose, face, and associated actions), and phonogram; (c) semagram (MOUTH), phonogram /r/ and phonetic complement (R).

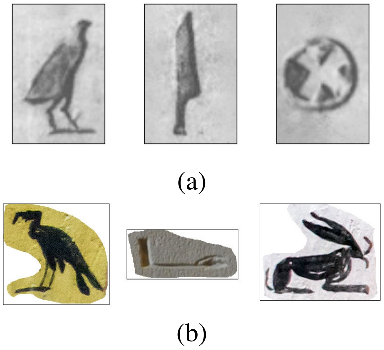


FIGURE 3. Sample images belonging to the D_1 (a) and D_2 (b) datasets.

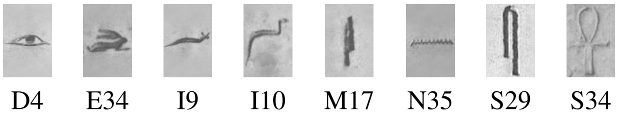


FIGURE 4. Examples of images labeled according to the Gardiner sign list.

contains labelled images of Egyptian hieroglyphs found in the Pyramid of Unas; the second one contains images selected and labelled by the authors, representing hieroglyphs on different supports. Fig. 3 shows some sample images belonging to the D_1 and D_2 datasets.

The hieroglyphs represented in the images are labelled according to the Gardiner sign list [22], i.e., with an alphabetic character followed by a number. Examples of image labelling is shown in Fig. 4.

1) ORIGINAL DATASETS

The first dataset [13], namely D_1, contains 4310 grayscale images representing hieroglyphs taken from photos of the walls inside the pyramid of Unas [23]. Each image has a dimension of 75 × 50 pixels and represents a single hieroglyph. The number of different hieroglyphs contained in this dataset is 172.

The second dataset, namely D_2, contains 1310 labelled color (RGB format) images, having variable dimensions and each representing a single element belonging to 48 possible different hieroglyphs. These images are taken from different documents, written by different hands and systems (sculpted, drawn, etc.) and belonging to diverse periods of Egyptian history, unlike those of the first dataset belonging to the same context. This choice increases the generalizability of the recognition. The images of the D_2 dataset were processed in order to uniform their dimensions and, for congruence with the D_1 dataset, the color information was converted to grayscale. As can be seen from Fig. 3-(b), the images were cut from scanned photos, with an irregular contour, and the final images were obtained by filling their rectangular bounding box with a white background. To achieve an uniform format, each image has been inscribed into a square, without altering the aspect ratio of the original data; then, the artificial white background has been removed and filled with a pseudorandom Gaussian texture resembling the original background around the hieroglyph. Eventually, the images have been scaled to obtain a dimension of 100 × 100 pixels. The preprocessing procedure described above is summarized in Fig. 5.

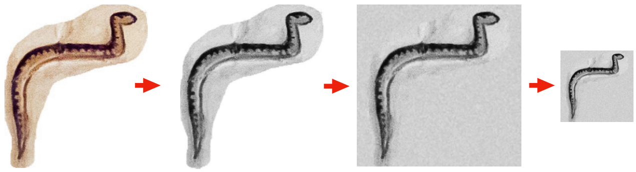


FIGURE 5. Preprocessing of images belonging to the D_2 dataset.

2) MERGED DATASET

The images in D_1 are characterized by the same physical support and appear quite uniform, whereas images in D_2 are taken from different supports (papyrus, stone, wood). Heterogeneous datasets are requierd by machine learning algorithms to achieve a high capacity of generalization. Therefore, the original D_1 and D_2 datasets were merged into a single dataset, referred to in the following as D. The joint dataset is not the union of D_1 and D_2, but only the images belonging to classes (i.e., hieroglyphs) contained in both datasets were selected. Let us denote the elements of the datasets as

$$D_1 = \{(x_i^{(1)}, y_i^{(1)})\} \quad \text{with} \quad i = 1, \dots, |D_1|$$
$$D_2 = \{(x_i^{(2)}, y_i^{(2)})\} \quad \text{with} \quad i = 1, \dots, |D_2|,$$

where $|D_k|$ denotes the cardinality of the kth dataset, $x_i^{(k)}$ an image within the kth dataset, and $y_i^{(k)}$ its label. Let Y_1 and Y_2 be the sets of the different labels in D_1 e D_2, respectively. Then, the final dataset is given by

$$D = \{(x_i, y_i)\} \quad \text{with} \quad y_i \in Y_1 \cap Y_2,$$

where (x_i, y_i) belongs either to D_1 or to D_2. At the end of this selection process, we achieved the final dataset D composed of 4309 images, distributed into 40 classes. The images belonging to the D_1 dataset have been extended to 75 × 75 (adding some background) and then resized to 100 × 100 pixels. Fig. 6 shows a histogram representing the number of images for each class in D. As can be seen, the dataset is quite unbalanced: this issue is addressed in the following section.

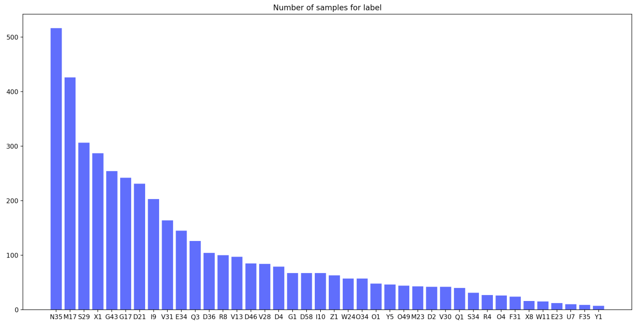


FIGURE 6. Number of images for each label (according to Gardiner sign list [22]) within the joint dataset *D*.

3) DATA AUGMENTATION AND DATASET BALANCING

In machine learning problems, a labelled dataset needs to be divided into training, validation and test subsets; in our experiments, the distribution of the entire dataset into such subsets was 70%, 15% and 15%, in that order. In order to limit the unbalanced dataset issue, data augmentation was applied to the training data. Augmentation is widely used in deep learning approaches to increase the size of the training set [24]: usually, it helps to increase the performance of the CNN and to reduce the overfitting problem. Augmentation consists in creating new labelled items from the available ones to be used for training. Transformations like translation, rotation and zooming, are often exploited: such operators are also helpful to make the network invariant to different orientations and scales.

As already mentioned, in this study, augmentation was used to achieve a better balancing of the training set. New items were added to classes containing a number of images below a threshold, whereas images were dropped from excessively numerous ones. More specifically, we used random translation along the horizontal and vertical directions (maximum 10 pixels), small random rotations (±10 degrees) as well as random zooming (zooming factor from 0.95 to 1.05). After this, since hieroglyphs can be oriented either left or right (without changing their meaning), augmentation also involved horizontal flipping. At the end, the maximum number of training images that a class can contain is set to two times the average number of images per label. The total number of images in the training set is 3014 and after data augmentation and downsampling, it becomes 3670, which increases to 7340 after flipping. In Fig. 7, the cardinality of each class of the augmented training dataset is shown.

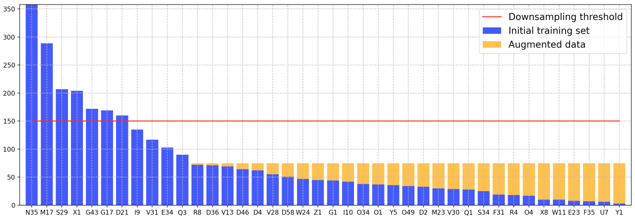


FIGURE 7. Final training dataset class population after data augmentation and downsampling.

III. ARCHITECTURES FOR HIEROGLYPH CLASSIFICATION

In this section, the architectures that were used in this study are introduced. First, some well-known state-of-the-art networks, used for image recognition tasks, namely ResNet-50, Inception-v3, and Xception, are recalled. All of them achieved excellent results on the ImageNet challenge. After this review, a new architecture, specifically designed for hieroglyphs classification, is proposed. All the architectures addressed here will be compared to each other in our experimental tests.

A. ARCHITECTURES FOR IMAGE CLASSIFICATION

ResNet [25], developed by Microsoft Research, introduced the Residual Network architecture in order to solve the problem of the vanishing/exploding gradient. In ResNet, the underlying mapping function that is looked for does not aim to map the input onto the output, but rather to model their difference, i.e., the residual. In other words, the underlying assumption is that it is easier optimizing the residual mapping than the original, unreferenced, mapping. Actually, the residual mapping approach is applied to successive blocks of the entire net, each composed by a few stacked layers, in which "shortcut" connection are used to compute the residual between the input and output of the block. In our tests, we used ResNet-50, which has over 23 million trainable parameters and was designed to process images with dimension 224 × 224 pixels.

Inception-v3 [26] is the third version of the GoogleNet model. The network has a depth of 42 layers; all convolution layers use the activation function ReLU and apply batch normalization (BN). After standard convolution and pooling layers to reduce the dimensions of the input, Inception-v3 presents its principal characteristic layers, i.e., the *inception modules*. They are parallel running convolution layers with different kernel sizes. The basic idea behind the inception modules is that they can extract similar features, but at different scales. The network, designed for the ImageNet challenge, processes images having a dimension of 299 × 299 pixels.

The Xception network [27] is an ''extreme'' evolution of the Inception model. In this network, the canonical inception block is simplified, making it similar to a depthwise separable convolution. This type of convolution consists of a spatial convolution performed independently over each channel of an input followed by a pointwise convolution; in the Xception module, instead, the order of the operations is 1 × 1 convolution first and then channel-wise spatial convolution. Another difference between Inception and Xception is the presence/absence of a non-linearity after the first operation. While in Inception model both operations are followed by a ReLU non-linearity, Xception does not introduce any non-linearity. As Inception-v3, the net was designed to process images with a dimension of 299 × 299 pixels.

B. PROPOSED NETWORK: GLYPHNET

In this section, a novel network, inspired by those previously described, is presented. The underlying idea is that, focusing on the specific task of hieroglyph recognition and tailoring the network on it, the complexity and overfitting issues can be reduced without, however, losing performance. In the next paragraphs, we will illustrate the network architecture and the choices of the hyperparameters for the training.

1) THE NETWORK ARCHITECTURE

At the basis of our model there are the separable convolutional layers, a feature also present in the Xception network. The overall organization of the network consists of six blocks, as detailed in the following.

1) The first block is the input stage and is composed of two identical sections, each composed by:
   a) a standard convolution layer, with 64 filters having a kernel size of 3 × 3, stride 1 × 1;
   b) a BN, a max pooling layer (with a kernel size of 3 × 3 and a stride of 2 × 2), and a ReLU.
2) Two identical blocks follow, each composed by:
   a) a separable convolutional layer, with 128 filters having a kernel size of 3 × 3, stride 1 × 1;
   b) a BN and a ReLU;
   c) a separable convolutional layer, with 128 filters having a kernel size of 3 × 3, stride 1 × 1;
   d) a BN, a max pooling layer (identical to the first block), and a ReLU.
3) Two blocks identical to the previous ones – with the only difference that the number of filters in the separable convolutional layers are 256 instead of 128 – follow.
4) The last block is the output stage and is composed of:
   a) a separable convolutional layer with 512 filters having a kernel size of 3 × 3, stride 1 × 1;
   b) a BN and a ReLU;
   c) a Global Average Pooling;
   d) a Dropout Regularizator with a fraction of the input units to drop equal to 0.15;
   e) a Fully Connected layer;
   f) a Softmax layer.

The architecture of the proposed network is sketched in Fig. 8. The dimensions of the input images has been set to 100 × 100 pixels, which represents a sufficiently compact size that allows the humans easily recognize the details of the hieroglyphic and the computation burden to be reduced without, nevertheless, affecting the classification performance.

Thanks to the reduction of the number of filters and of the number of layers with respect to the architectures described in Section III-A, the proposed network has a much lower number of parameters, which is only 498856 (of which 494504 are trainable), compared to the over 20 million parameters of the classical networks.

2) NETWORK HYPERPARAMETERS

The proposed architecture was implemented and trained using Keras [28] and TensorFlow [29], the well-known open-source software libraries providing a Python interface to artificial neural networks and machine learning systems design.

The loss function used to optimize the model during training is the categorical cross-entropy. As optimization method, ADAM (*adaptive moment estimation*) [30] was used, with a batch size of 32 images.

To improve the generalization capacity of the network, we used batch normalization after each convolution, $L2$-Loss to regularize the weights in the fully connected layer of the final block as well as a dropout layer. An adaptive learning rate was also used: the initial learning rate was set to 0.001 and halved every 15 epochs. Initial learning rate and decay factor were chosen empirically, after having performed several tests with a trial-and-error strategy.

All weights in dense, convolutional and separable convolutional layers were inizialized with the Glorot Uniform strategy without bias.

C. TRANSFER LEARNING

The networks described in Sections III-A can be trained on a given image dataset either ''from scratch'' or using the *transfer learning* [31] approach. The idea of transfer learning is to use previous knowledge acquired for one task to solve newer and related ones. In transfer learning, the weights/filters learned by a known architecture – trained on a given task by means of a huge dataset – are reused to face a different task. The underlying idea here is that early stages are devoted to extract basic image descriptors [32], while the final ones are dedicated to combine them for the specific classification problem we are dealing with.

In this study, we used the pre-trained models described in Section III-A, exploiting most of the bottom layers of the original architectures with frozen weights, and substituting the last output layers with new fully connected layers to be trained with our dataset. As an example, the new top layers applied to the ResNet-50 were chosen as follows:

- a Global Average Pooling layer;
- a Dense layer with 128 units;

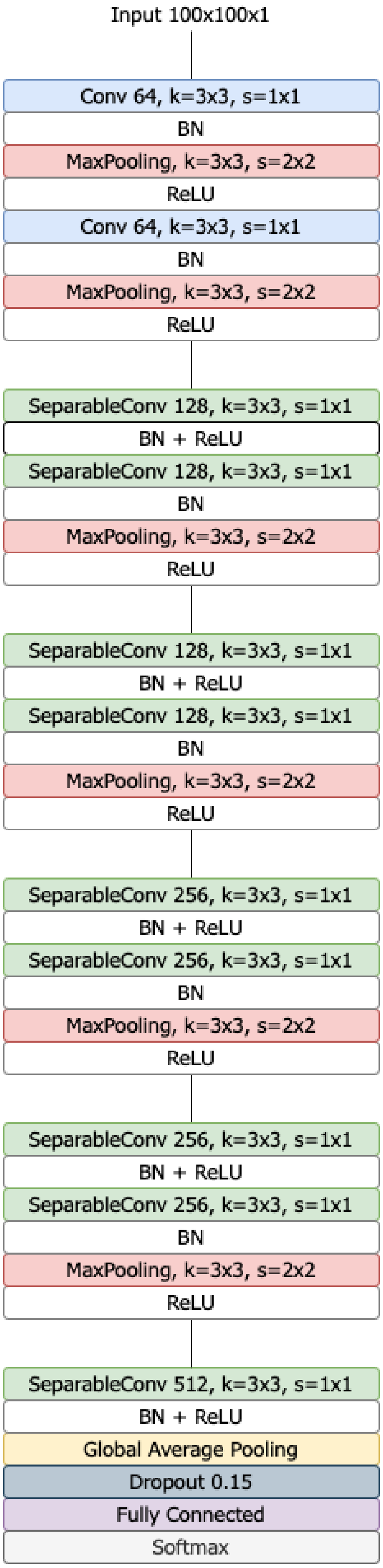


FIGURE 8. The proposed Glyphnet architecture.

- a Dropout layer with dropout rate equal to 0.15;
- a Dense layer with 40 neurons (number of classes);
- a Softmax activation layer as output for classification.

Similar layers, with small changes, were used for the Inception-v3 and Xception networks. The performance obtained from transfer learning will be shown for comparisons in the next section.

IV. EXPERIMENTAL RESULTS

In this section, the performance of the classical networks described in Section III-A and the new Glyphnet, proposed in Section III-B, are compared to each other. The evaluation metrics are the standard *accuracy*, *precision*, *F1-Score*, and *recall*, which are commonly used to evaluate classification performances. The performance of all the networks were obtained by using the hieroglyph dataset *D* described in Section II-B, with data augmentation applied to the training set. Computation times, relative to training and prediction, were also evaluated for the different architectures.

Table 1 reports the evaluation metrics obtained by using the transfer learning approach described in Section III-C applied to the standard networks described in Section III-A, i.e., ResNet-50 [25], Inception-v3 [26], and Xception [27], whereas Table 2 reports the same metrics when the different networks, including Glyphnet, were trained from scratch. Comparing Tables 1 and 2, we observe that the classical networks yield better results when they are trained from scratch with respect to the case when transfer learning is used. The results in Table 2 show also that the new architecture proposed in this study yields, with reference to all the metrics, the best performances.

TABLE 1. Networks performance by using the transfer learning approach.

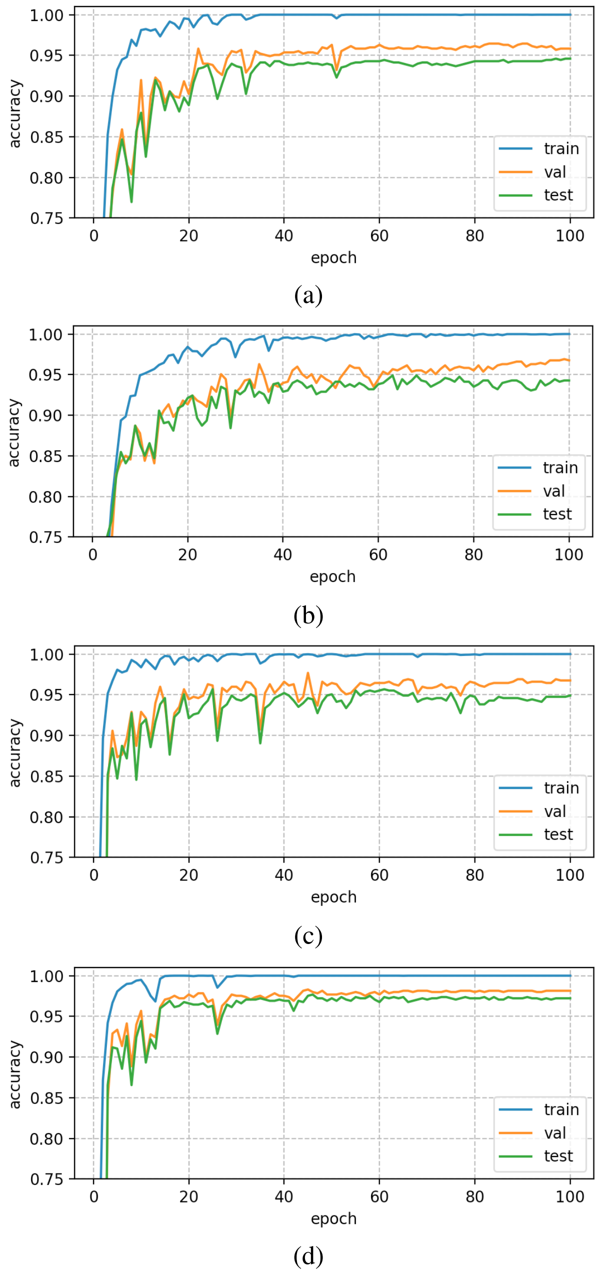
Architecture	Accuracy	Precision	Recall	F1-Score
ResNet-50	0.906	0.882	0.825	0.840
Inception-v3	0.864	0.730	0.737	0.717
Xception	0.834	0.715	0.720	0.703

TABLE 2. Networks performance by using training from scratch.

Architecture	Accuracy	Precision	Recall	F1-Score
ResNet-50	0.945	0.919	0.905	0.903
Inception-v3	0.948	0.917	0.904	0.900
Xception	0.956	0.919	0.930	0.919
Glyphnet	**0.976**	**0.975**	**0.965**	**0.968**

In order to better understand the robustness of the networks and achieve useful insights on their training process, Figs. 9 and 10 show, for all the models that were analysed, the trend of the accuracy and of the loss function, respectively, vs. the epoch time during the training, validation and testing processes.

In order to show the robustness of our tests, the results of a stratified 5-fold cross-validation over the entire dataset for the Glyphnet are shown in Fig. 11. Shaded areas enclose the curves obtained at each run of the cross-validation. As can be seen, the mean of the curves vs. the epoch time tends to be flat and the standard deviation tends to zero.



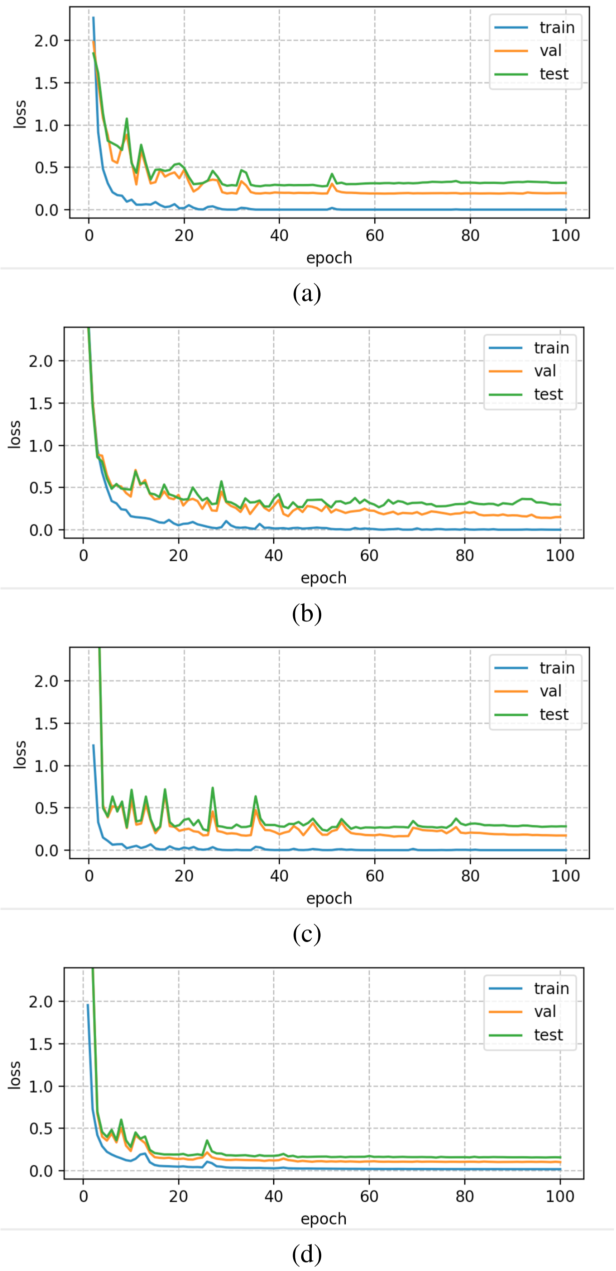
(a)

(b)

(c)

(d)

FIGURE 9. Evolution of the accuracy function vs. epoch time during the training, validation and testing processes for the different architectures: (a) ResNet-50; (b) Inception-v3; (c) Xception; (d) the proposed Glyphnet.



(a)

(b)

(c)

(d)

FIGURE 10. Evolution of the loss function vs. epoch time during the training, validation and testing processes for the different architectures: (a) ResNet-50; (b) Inception-v3; (c) Xception; (d) the proposed Glyphnet.

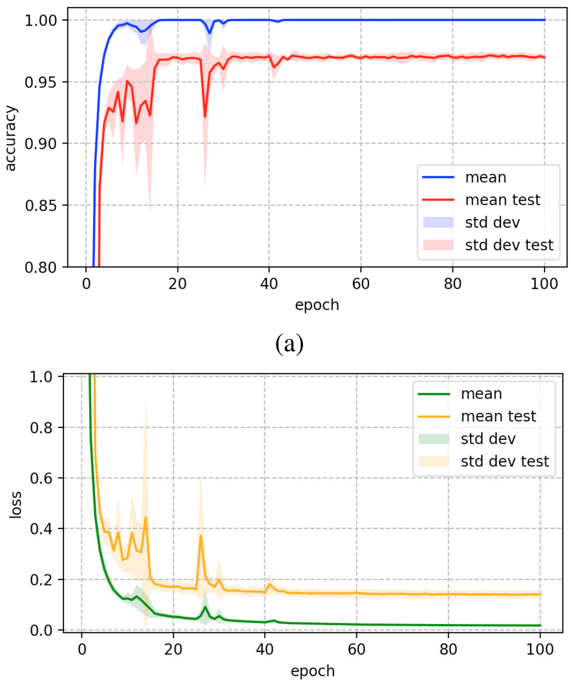
Some results were also obtained by eliminating the possibility of data augmentation within the training set, in order to better understand the importance of its introduction. With reference to the proposed Glyphnet, Fig. 12 shows the evolution of accuracy vs. the epoch time when data augmentation is either used or not in the training set. As can be seen, the beneficial effects of data augmentation is remarkable.

In order to evaluate the computational burden of the analysed networks, we report in Fig. 13 the training time, measured as milliseconds per training step (gradient updates), as a function of the image dimension and for a batch size equal to 32. Times have been measured on an NVIDIA Tesla T4 GPU. As can be noticed, the proposed network is the fastest among all tested CNNs, for all image dimensions.

The times related to the prediction of the images were also calculated on the same GPU. Batch-size was set to 1 and 32. The results, obtained averaging 100 measures, are reported in Table 3. As can be seen, also in this case, the proposed Glyphnet is the fastest in the task of image prediction for any image dimensions.

TABLE 3. **Prediction runtimes for different batch-sizes and different image dimensions.**

	Batch-size=1			Batch-size=32		
Architecture	100x100	200x200	300x300	100x100	200x200	300x300
ResNet-50	8	10	13	29	86	181
Xception	7	10	11	27	110	275
Inception-v3	12	13	14	20	58	116
Glyphnet	**3**	**3**	**4**	**10**	**22**	**50**



(a)

(b)

FIGURE 11. **Results of 5-fold cross validation for the proposed Glyphnet: accuracy (a) and loss (b) vs. epoch time.**

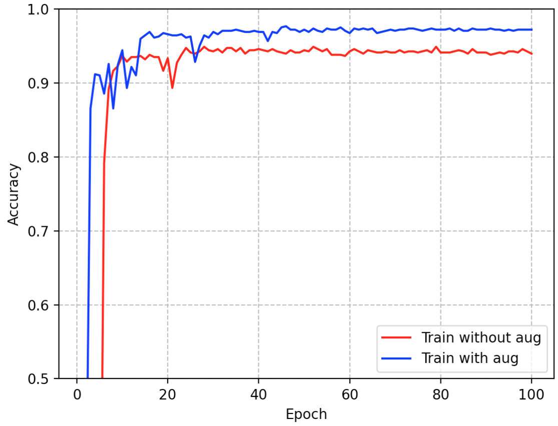


FIGURE 12. **Effect of data augmentation on the training set.**

Summarizing, our proposed architecture, Glyphnet, has shown the best performances in terms of classification rate and computational time. This is related to a simpler architecture, with fewer parameters to be trained and, thus, less prone to overfitting.

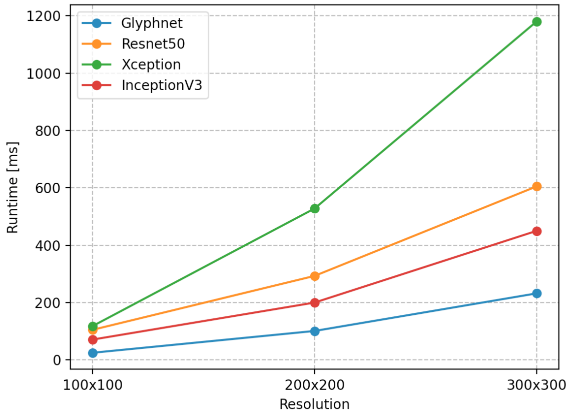


FIGURE 13. **Training computation time.**

V. CONCLUSION

In this work, we have explored the capability of deep learning techniques to face the problem of ancient Egyptian hieroglyphs classification. To this aim, standard networks already proposed to tackle image recognition tasks have been tested and a new one, Glyphnet, has been developed and tailored for the specific purpose at hand. Two image datasets of labelled hieroglyphs were used to train and test the networks. Performances were measured using standard metrics, giving significant results for all the tested networks, with the proposed Glyphnet outperforming the others, in terms of performance as well as ease of training and computational saving.

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. Actually, there are several open issues that may benefit from the use of the proposed approaches: coding, recognition and transliteration of hieroglyphic signs; recognition of determinatives and their semantic field; toposyntax of the hieroglyphic signs combined to form words; linguistics analysis of the hieroglyphic texts; recognition of corrupt, rewritten, and erased signs, towards even the identification of the ‘‘hand’’ of the scribe or the school of the sculptor.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[2] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, Dec. 2019, Art. no. 045002, doi: 10.1103/RevModPhys.91.045002.

[3] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and M. Prabhat, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, pp. 195–204, Feb. 2019.

[4] K. Yu, A. Beam, and I. Kohane, "Artificial intelligence in healthcare," *Nature Biomed. Eng.*, vol. 2, no. 10, pp. 719–731, 2018.

[5] C. Scapicchio, M. Gabelloni, A. Barucci, D. Cioni, L. Saba, and E. Neri, "A deep look into radiomics," *La Radiol. Medica*, vol. 4, pp. 1–16, Jul. 2021.

[6] A. Barucci and E. Neri, "Adversarial radiomics: The rising of potential risks in medical imaging from adversarial learning," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 47, no. 13, pp. 2941–2943, Dec. 2020, doi: 10.1007/s00259-020-04879-8.

[7] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," *Int. J. Hum.-Comput. Interact.*, vol. 36, no. 6, pp. 495–504, Apr. 2020.

[8] R. Chatila, V. Dignum, M. Fisher, F. Giannotti, K. Morik, S. Russell, and K. Yeung, "Trustworthy AI," in *Reflections on Artificial Intelligence for Humanity* (Lecture Notes in Computer Science), vol. 12600, B. Braunschweig and M. Ghallab, Eds. Cham, Switzerland: Springer, 2021, pp. 13–39, doi: 10.1007/978-3-030-69128-8_2.

[9] T. Clanuwat, A. Lamb, and A. Kitamoto, "KuroNet: Pre-modern Japanese Kuzushiji character recognition with deep learning," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sydney, NSW, Australia, Sep. 2019, pp. 607–614.

[10] A. Lamb, T. Clanuwat, and A. Kitamoto, "KuroNet: Regularized residual U-Nets for end-to-end Kuzushiji character recognition," *Social Netw. Comput. Sci.*, vol. 1, no. 3, May 2020, Art. no. 177.

[11] E. Roman-Rangel and S. Marchand-Maillet, "Indexing mayan hieroglyphs with neural codes," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 253–258.

[12] F. de Saussure, *Course in General Linguistics*, P. Meisel and H. Saussy, Eds. New York, NY, USA: Columbia Univ. Press, 2011.

[13] M. Franken and J. van Gemert, "Automatic Egyptian hieroglyph recognition by retrieving images as texts," in *Proc. 21st ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, 2013, pp. 765–768, doi: 10.1145/2502081.2502199.

[14] M. Franken. (2017). *Glyphreader*. [Online]. Available: https://github.com/morrisfranken/glyphreader

[15] J. Duque-Domingo, P. Herrera, E. Valero, and C. Cerrada, "Deciphering Egyptian hieroglyphs: Towards a new strategy for navigation in museums," *Sensors*, vol. 17, no. 3, p. 589, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/3/589

[16] R. Elnabawy, R. Elias, and M. Salem, "Image based hieroglyphic character recognition," in *Proc. 14th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2018, pp. 32–39.

[17] E. Iglesias-Franjo and J. Vilares, "TIR over Egyptian hieroglyphs," in *Proc. 27th Int. Workshop Database Expert Syst. Appl. (DEXA)*, 2016, pp. 198–203.

[18] S. Rosmorduc, "Automated transliteration of Egyptian Hieroglyphs," in *Meeting of the Computer Working Group of the International Association of Egyptologists*. Piscataway, NJ, USA: Gorgias Press, Jul. 2008, pp. 167–183.

[19] M. Nederhof, "OCR of handwritten transcriptions of Ancient Egyptian hieroglyphic text," in *Proc. Altertumswissenschaften Digit. Age, Egyptol., Papyrol. Beyond*, M. Berti, Eds., Leipzig, Germany, Nov. 2015.

[20] *Unravel the Symbols of Ancient Egypt*. Accessed: Jul. 26, 2021. [Online]. Available: https://blog.google/outreach-initiatives/arts-culture/unravel-symbols-ancient-egypt

[21] *FABRICIUS*. Accessed: Jul. 26, 2021. [Online]. Available: https://artsexperiments.withgoogle.com/fabricius/en

[22] A. Gardiner, *Egyptian Grammar*. Oxford, U.K.: Griffith Institute, 1957.

[23] A. Piankoff, *The Pyramid UNAs* (Bollingen Series). Princeton, NJ, USA: Princeton Univ. Press, 1969.

[24] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, 2019.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[27] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[28] F. Chollet (2015) *Keras*. [Online]. Available: https://github.com/fchollet/keras

[29] A. Agarwal. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: http://tensorflow.org/

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[31] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, vol. 2. Cambridge, MA, USA: MIT Press, 2014, pp. 3320–3328.

[32] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1717–1724.

ANDREA BARUCCI received the Laurea degree *(cum laude)* in physics, the Ph.D. degree in electronic systems engineering, and the Medical Physical Expert degree from the University of Florence, Florence, Italy, in 2006, 2011, and 2015, respectively. Since 2010, he has been with the Institute of Applied Physics "Nello Carrara" (IFAC), National Research Council (CNR), where he was initially a Postdoctoral Researcher and then has been a Researcher, since 2018. Since 2017, he has been an Adjunct Professor with the Department of Biomedical, Experimental and Clinical Sciences "Mario Serio," University of Florence. He has been involved in several research projects in the fields of radiomics and artificial intelligence focusing on clinical image processing and precision medicine. His research interests include medical physics, clinical imaging, biobanks, precision medicine, photonics, biophotonics, machine learning, deep learning, and artificial intelligence.

COSTANZA CUCCI received the Laurea degree in physics and the Ph.D. degree in conservation science from the University of Florence, Italy, in 1996 and 2004, respectively. She is currently a Researcher with the Institute of Applied Physics "Nello Carrara" (IFAC), Italian National Research Council (CNR), Florence, Italy. Since 2000, she has been carrying her research activity within CNR in the field of applied spectroscopy, with a special focus on testing protocols for non-invasive methodologies, cultural heritage, and environmental monitoring. Her current research interests include hyperspectral imaging applied to cultural heritage, multivariate and statistical data analysis, multimodal non-invasive methods for *in-situ* analysis of materials and their degradation processes in artworks, and spectroscopy for sensor applications. She is a Coordinator of the "Indoor Lighting of Cultural Heritage" Working Group of the Italian Standardization Body (UNI) and Expert Member of the European Committee for Standardization (CEN) Technical Committee "Exhibition Lighting of Cultural Property."

MASSIMILIANO FRANCI is currently an Associate Professor of Egyptology and anthropology of food with the CAMNES-LdM Institute. He has been involved in many projects on public archaeology. His monograph on Ancient Egyptian Astronomy has been translated into Arabic and officially presented with Cairo University, Egypt. His research interests include ancient Egyptian linguistics and philology, Egyptian cultural identity, cultural relationships between Egypt and the ancient near east, and the history of Deir el Medina Village. He is a member of the International Association of Egyptologists, the Italian Society of History of Religions, and the International Association of History of Religion. He is a Representative Member of CAMNES in the cooperation agreement with the Institute of Mediterranean and Oriental Cultures, Polish Academy of Sciences, Poland.

MARCO LOSCHIAVO received the B.Sc. and M.Sc. degrees in computer sciences, in 2018 and 2021, respectively. His research interests include deep learning, machine learning, image processing, and classification.

FABRIZIO ARGENTI (Senior Member, IEEE) received the Laurea degree *(cum laude)* in electronics engineering and the Ph.D. degree in electronics and information engineering from the University of Florence, Florence, Italy, in 1989 and 1993, respectively. Since 1993, he has been with the Department of Electronics and Telecommunications (now Department of Information Engineering), University of Florence, where he was initially an Assistant Professor and then has been an Associate Professor, since 2002. His teaching experience includes courses on digital signal processing, estimation theory, and information theory. He has been involved in several research projects in the fields of image processing, multimedia transmission, and remote sensing. His research interests include image processing, multiresolution analysis, statistical signal processing, remote sensing, inverse problems, and machine learning approaches to signal processing.