Dua-Khety:
Hieroglyphics Detection, Classification and Transliteration

The American University in Cairo - Department of Computer Science and Computer Engineering (CSCE)

Ahmed Agha, Malak Sadek, Mohamed Badreldin, Mohamed Ghoneim

Supervised by Dr. Amr Goneid

20th of May, 2018

Contents

Abstract

This paper introduces Dua-Khety; a mobile application available on both Android and iOS devices that is capable of localizing and segmenting hieroglyphs within images and then classifying them and assigning them a Gardiner Code to provide the user with their transliteration and meaning. The entire process takes place offline, meaning that the service is convenient, fast, lightweight, and can be accessed anywhere at anytime. The application also has other features, such as the ability to browse other users' photographs, and access a hieroglyphics dictionary, to provide an engaging and educative experience.

Keywords: Machine Learning, Computer Vision, Mobile Apps, Hieroglyphs, Egyptology

1 Introduction

Egyptian hieroglyphs are one of the most illusive pictographic languages in the world. Pictographic languages do not form full and proper sentences, rather they use a series of pictures and symbols to convey the semantics of the sentence. This can sometimes be literal (i.e. a symbol of a man refers to a man) and other times metaphorical (i.e. some symbols of animals refer to certain Egyptian gods) [1]. Hieroglyphs also present an additional challenge in that there is no standardized reading order as they can be written top to bottom, left to right, or right to left, and sometimes in multiple orientations with several overlaps between the symbols.

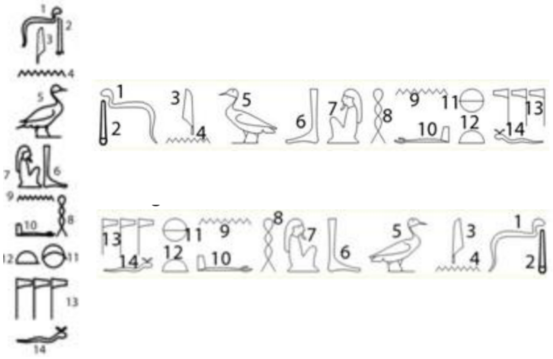
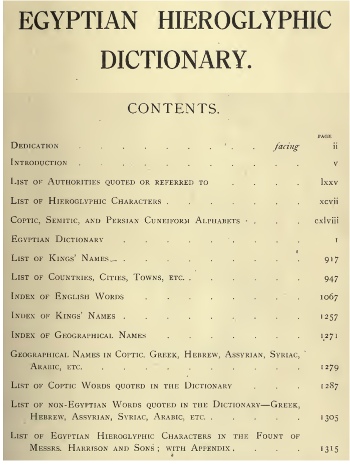


Figure 1: Examples of hieroglyphs read in different orders [2]

What's more is that over the years, most of the temples and documents where these hieroglyphics can be found have become eroded, worn down, fractured, or vandalized, making any digital visual processing all the more difficult. Because of these obstacles, much of the way of life of the ancient Egyptians has been lost as no one was able to properly decipher the meanings behind most of their symbols. Even though some advancements were made [3], they could not truly unlock the key to Egypt's past. Eventually, Sir Alan Gardiner documented and classified each hieroglyphic symbol with a code that could then be used to search for more information about it's meaning. This hieroglyphic dictionary ended up being thousands of pages long and searching for symbols became a tedious task, especially as one had to know their code to search for them. Even with the emergence of the internet speeding up manual searches, there still remained the problem of not knowing a hieroglyphic's 'Gardiner Code' to be able to search for it.



EGYPTIAN HIEROGLYPHIC DICTIONARY.

CONTENTS.

		PAGE
Dedication	*facing*	ii
Introduction		v
List of Authorities quoted or referred to		lxxv
List of Hieroglyphic Characters		xcvii
Coptic, Semitic, and Persian Cuneiform Alphabets		cxlviii
Egyptian Dictionary		1
List of Kings' Names		917
List of Countries, Cities, Towns, etc.		947
Index of English Words		1067
Index of Kings' Names		1257
Index of Geographical Names		1271
Geographical Names in Coptic, Greek, Hebrew, Assyrian, Syriac, Arabic, etc.		1279
List of Coptic Words quoted in the Dictionary		1287
List of non-Egyptian Words quoted in the Dictionary—Greek, Hebrew, Assyrian, Syriac, Arabic, etc.		1305
List of Egyptian Hieroglyphic Characters in the Fount of Messrs. Harrison and Sons; with Appendix		1315

Figure 2: The contents page for Sir Alan Gardiner's massive hieroglyphic dictionary [4]

To help facilitate the decoding of hieroglyphs and extending that knowledge beyond archaeologists and field experts and into the hands of the general public, Dua-Khety was created. In ancient Egypt, only very few of the elite citizens could read and write. As a result, scribes used to read and write letters for common people and it was considered a very prestigious profession. One scribe named Dua-Khety famously wrote a letter to his son titled The Satire of the Trades [5] where he dismisses other professions and talks about the beauty of being a scribe saying:

> "So I would have you love writing more than your mother and have you recognize its beauty. For it is greater than any profession, and there is none like it on earth."

The project honors the importance of Egyptian scribes in preserving their people's knowledge and way of life for centuries by naming the application in Dua-Khety's name with the aim of essentially placing a digital Egyptian scribe in people's pockets.

2 Related Work

There are currently no widely accepted models for hieroglyph detection and classification, and none of them are implemented on mobile devices. The most comprehensive work done is by Morris Franken and Jan Gemert [3] who provide a extensive dataset of hieroglyphic symbols organized by their Gardner Codes, which is utilized in this project. Their method entails essentially treating the symbols as text to localize individual symbols and using different methods such as the bag-of words (BOG) approach [6] and a combination of HOG [7] and Shape-Context [8] techniques to classify them. Their dataset is also the most realistic when compared to the noisy and irregular real-life context that hieroglyphs are found in, as opposed to other papers who manually extract and clean individual hieroglyphs [7–9]. Similar work was done by Mark-Jan Nederhof at the University of St Andrews [10] who uses optical character recognition (OCR) on handwritten transcriptions of hieroglyphs, essentially treating tehm as text as well, instead of using the hieroglyphs in their natural form. While there is a large amount of work done on Western and Asian languages [11–14], what little work exists on pictographic languages is usually focused on Meso-American and Mayan languages as opposed to Egyptian hieroglyphs [6–9]. These papers also mainly rely on computer-vision techniques such as Gradient Field Histograms and Histograms of Orientation Shape Context (HOSC).

The project thus presents a new approach to localizing and detecting hieroglyphics, using a mixture of computer vision and machine learning techniques by combining pre-processing and edge detection with a neural network that is then used to classify the hieroglyphics. The entire process occurs offline on a mobile device, introducing a truly novel application.

3 Dataset

The dataset used was provided by Morris Franken and Jan C. van Gemert from the University of Amsterdam [3]. It contains 4,210 manually annotated images of Egyptian hieroglyphs found in the Pyramid of Unas in Egypt. Within the dataset there are around 200 different classes (e.g. humans, flowers, etc.) with the number of images per class varying significantly from 1 to 300 images. This was undesirable for the classification step as the dataset was severely unbalanced due to the varying frequency of appearance across different symbols. To overcome this, only the 22 largest classes were used. The number of images per class still varied but capped at the minimum of around 10 pictures.

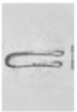
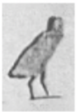
Image			
Gardener Label	S29	V13	G43

Figure 4: Example symbols from the dataset

4 Implementation

4.1 Pre-Processing and Segmentation

OpenCV was used to to pre-process and segment the images taken by users, isolating individual hieroglyphs. Firstly the images are converted to gray-scale and blurred with a radius of three, which was found to have a better result at reducing noise than various types of filters. Following this, the image pixel average is computed and it is thresholded with a minimum value of the computed average and a maximum value of 255. Afterwards, Canny Edge Detection is applied with thresholds of:

$$0.66 * imageAverage$$

and

$$1.33 * imageAverage$$

as well as an aperture size of three. These values were found to be optimal through experimentation to balance between removing edges that cause noise and avoiding rendering the shapes indistinguishable.

The image is then diluted and eroded, and is then segmented based on vertical columns or horizontal rows to be able to determine the orientation of the hieroglyphics by using Hough Linear Transform to detect the largest lines (which separate rows or columns).

The final step is to use a Connected Components List to create bounding boxes around individual hieroglyphics based on the determined order. These bounding boxes are placed on the original image taken and used to crop out and extract the individual symbols for classification and presentation to the user. These cropped symbols from the original image are only turned into black and white and thresholded in the same way (without the Canny Edge Detection as this significantly lowers the classification accuracy).

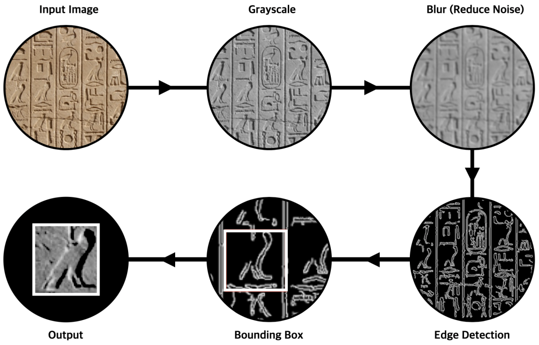


Figure 6: A summary of the pre-processing and segmentation steps

4.2 Classification

As a first step the Xception network was used, however, it appeared that different classes needed different weights matrices as some classes required classifiers that pay close attention to details, while others only needed to consider the outer shape. For example, class G symbols (which represent birds) achieved a 69% accuracy with Xception, as it needed to search for small details to differentiate classes. On the other hand, class N symbols (which represent elements like the sun moon, and sea) have much less details and did not have a good validation accuracy with Xception. Those symbols reached 65% with a CNN of 2 layers however (using maxpooling, followed by dropout, then flattening the output, and using 1 dense layer of 128 neurons, followed by another dropout, and a final dense later who's neurons are the same number of classes to be distinguished). As there is no way of knowing which class a hieroglyphic in a picture taken by the user is in, CNNs were used instead of Xception.

The initial classifier used consists of a 32 matrix convolutional layer with a kernel size of 3x3 and Relu as an activation function followed by an identical 64 matrix convolutional layer. This is followed by max pooling with a pool size of 2x2 and a drop out of 0.25. The model is then flattened and this is followed by a dense layer 40 with a Relu activation function that is also flattened. Finally, a dense layer 22 representing the 22 classes that sir Gardiner has used to categorize hieroglyphics is used. This layer's activation function is Soft Max and it uses an RMP Prop optimizer.

32 Matrix Convolutional Layer	
Kernal Size	3x3
Activation Function	Relu
64 Matrix Convolutional Layer	
Kernal Size	3x3
Activation Function	Relu
Max Pooling	
Pool Size	2x2
Drop Out	
Value	0.25
Flatten Model	
Dense Layer 40 (Flattened)	
Activation Function	Relu
Drop Out	
Value 0.5	
Dense Layer (22 vs. 162)	
Activation Function	Soft Max
Optimizer	RMS Prop

Figure 7: Summary of the initial machine learning model

The resulting .h5 file was optimized and frozen to be converted into a .pb file that is then used within the Tensorflow framework on Android devices, and is also fed into the CoreML conversion tools to fit within the CoreML framework on iOS devices. The classifier accepts images of dimensions 150x150 (any incoming images from the application are resized to these dimensions). Initially, classifier results were poor, however, they swiftly improved after binarizing the images before feeding them to the classifier. While the following results were achieving around 70% accuracy, it was found that the trained model was over-fitting on both the training and testing data. After looking for the root of the problem, it was found to be the dataset which was extracted from one place (the Pyramid of Unas) in one go from one person using one device and was in fact captured from a book of images. The testing data was simply not diverse enough as the aim of the system is to classify hieroglyphs in different temples and on different stones.

Moving forward, hieroglyphic scriptures from the Louvre, the British Museum, and the Museum of Alexandria were obtained as new testing data, as they proved to be more diverse. In addition to this, the previous dataset from the Pyramid of Unas was also augmented and cleaned to remove misleading symbols. Augmentation used a shear range of 0.2, zoom range of 0.2, and rotation range of 0.2.

With some research, it was found that Siamese Neural Networks would be more suitable to solve the problem at hand. These networks differ from normal ones as they take pairs of images as an input, as well as a label representing whether those images are from the same class or not (displayed as a 0 or 1). In other words, half of what is fed to the network are pairs of images of the same classes with their label as 1, and the other half, are two different images from two different classes, with their label as 0. The images are chosen randomly. The model consists of four separable convolutional layers with varying kernel sizes and 'relu' as an activation function with Max Pooling between each one. The model is flattened at the end to get the features at that stage. For training, two of these exact layers are merged and another dense layer is added at the end with one output to represent whether the pairs of images are from the same class or not, that uses 'sigmoid' as an activation function.

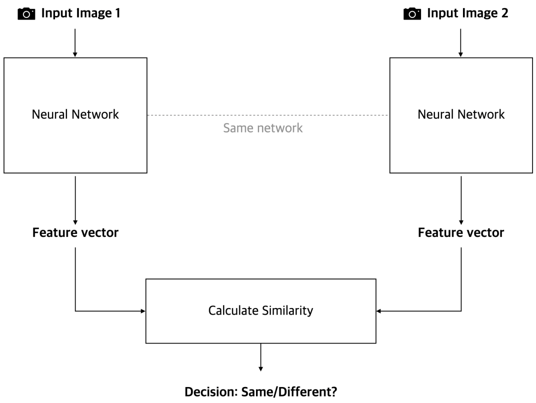


Figure 9: The architecture of the Siamese Network used

4.3 The Mobile Applications

Applications were created for both Android devices (using Java) and iOS devices (using Swift and Objective-C). The appropriate OpenCV libraries were used for both applications and CoreML was used to integrate the machine learning model into the iOS application, while Tensorflow was used for the Android application. The iOS version of the application was uploaded to Apple's App Store (view here), while the Android version was uploaded to Google's Play store (view here).

The applications also offer a number of other features. Users can look at their own search history or browse a social feed showing other users' recent photographs (which they can then analyze on their own device). A symbol dictionary provides users with a digitized version of Sir Gardner's dictionary which they can browse through or use to search for specific symbols. Users can also send new pictures they take to the development team in order to gradually increase the size of the dataset thus improving the quality of the classifier through crowd-sourcing. All the features except for the social feed fully work offline. This was an important goal for the project as most temples, tombs, etc. are out in the desert where there is no proper cellular connection or WiFi. Firebase was used for storing the database required to show the symbol dictionary, as well as user submissions for their history and social feed. It was also used for account management (which was optional for transliterating, but necessary to access the social feed).

Finally, after taking a picture, the application presents the user with a list of the hieroglyphics found within it. When the user selects one symbol, they can see its Gardner code, its literal meaning, its semantic meaning, and links for more information. This is considered a transliteration as opposed to a translation as hieroglyphics are not grammar-based but rather work by stringing together words to form larger semantic contexts (as mentioned earlier in the paper) [1]. Screenshots of the iOS application's user interface can be found in Appendix A.

5 Results

Classification began at 2% accuracy when using a deep layer network with 100+ layers, and using the raw images provided by Franken and Gemet. Binarizing the images increased the accuracy to 10% as the dataset became gray-scale but the actual images taken from the application were still colored. Augmenting the images in the dataset increased accuracy by a further 3%. Using a smaller neural network increased the accuracy to 55% and increasing the kernel size to 30x30 reached an accuracy of 70% however it suffered from over-fitting. The final classification accuracy reached with Siamese Networks was a **66% accuracy when predicting the top match**, and an **82% accuracy when predicting the top five matches.**

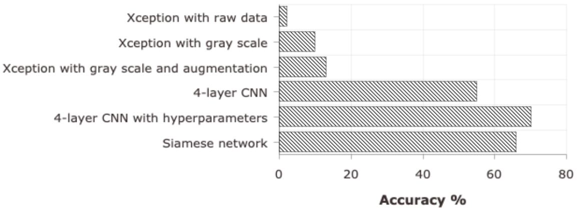


Figure 11: A graph of the classifier's accuracy as the project progressed

6 Conclusion

Dua-Khety has been presented as an iOS and Android mobile application capable of detecting, segmenting, and classifying hieroglyphics completely offline. Providing other features also makes the applications more comprehensive and engaging. By utilizing computer-vision pre-processing and segmentation techniques, as well as a Siamese Neural Network for classification, the application presents a novel way of solving a mystery as old as civilization itself.

References

[1] M. Halliday, "Spoken and written language," *Deakin University Press, p.19*, 1985.

[2] B. d'Alexandría, "Hieroglyphs step by step." https://www.bibalex.org/learnhieroglyphs/lesson/LessonDetails_En.aspx?l=57. Accessed: 2018-02-30.

[3] M. Franken and J. C. van Gemert, "Automatic egyptian hieroglyph recognition by retrieving images as texts," in *MM '13*, 2013.

[4] A. Gardiner, *Egyptian Hieroglyphic Dictionary.* John Murray, London, 1920.

[5] J. Mark, "The satire of trades." https://www.ancient.eu/article/1074/the-satire-of-the-trades/. Accessed: 2018-03-02.

[6] G. C. O. J. G.-P. D. Roman-Rangel, E., "Searching the past: an improved shape descriptor to retrieve maya hieroglyphs," *ACM MM*, 2011.

[7] T. B. Dalal, N., "Histograms of oriented gradients for human detection," *INCVPR*, 2005.

[8] M. J. Belongie, S., "Shape matching and object recognition using shape contexts," *TPAMI, 24(4)*, 2002.

[9] O. B. E. Frauela Y., Quesadab, "Detection of a mesoamerican symbol using a rule-based approach.," *Pattern Recognition, 39*, 2006.

[10] M.-J. Nederhof, "Ocr of handwritten transcriptions of ancient egyptian hieroglyphic text.," *In Altertumswissenschaften in a Digital Age: Egyptology, Papyrology and beyond, proceedings of a conference and workshop in Leipzig*, 2015.

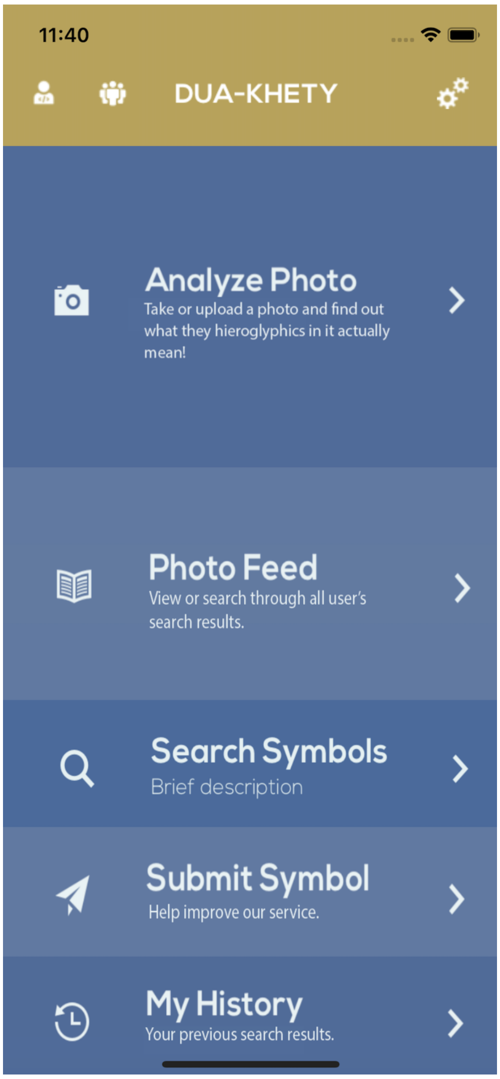
[11] O. E. W. Y. Epshtein, B., "Detecting text in natural scenes with stroke width transform.," *InCVPR, 2963-2970*, 2010.

[12] G. J. G. T. Karaoglu, S., "Object reading: Text recognition for object recognition," *InECCV-IFCVCR Workshop*, 2012.

[13] G. J. G. T. Karaoglu, S., "Con-text:text detection using background connectivity for fine-grained object classification," *ACM-MM*, 2013.

[14] T. K. T. S. M. Y. Kimura, F., "Modified quadratic discriminant functions and the application to chinese character recognition.," *PAMI*, 1987.

Appendix A iOS Application Screenshots


11:40

DUA-KHETY

Analyze Photo
Take or upload a photo and find out what they hieroglyphics in it actually mean!

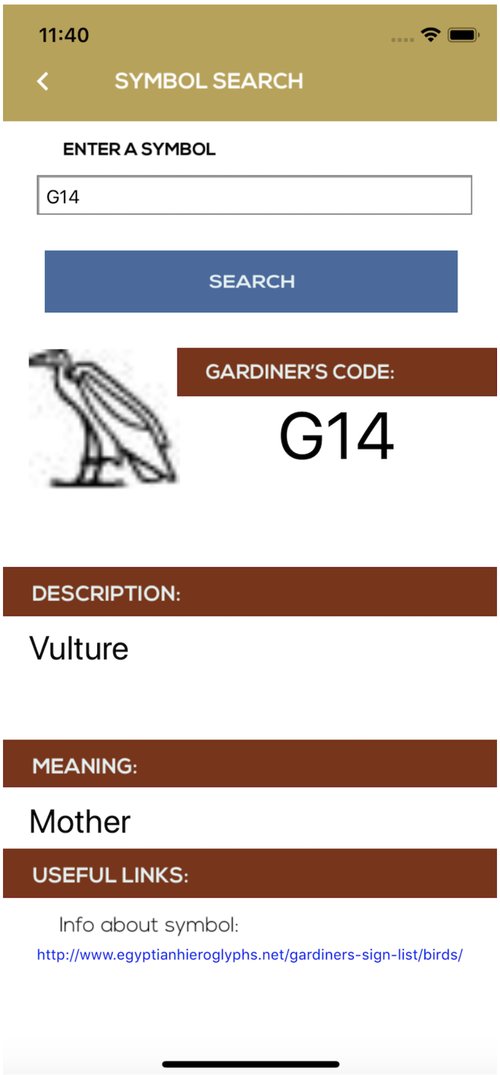
Photo Feed
View or search through all user's search results.

Search Symbols
Brief description

Submit Symbol
Help improve our service.

My History
Your previous search results.

(a) Main Menu


11:40

SYMBOL SEARCH

ENTER A SYMBOL

G14

SEARCH

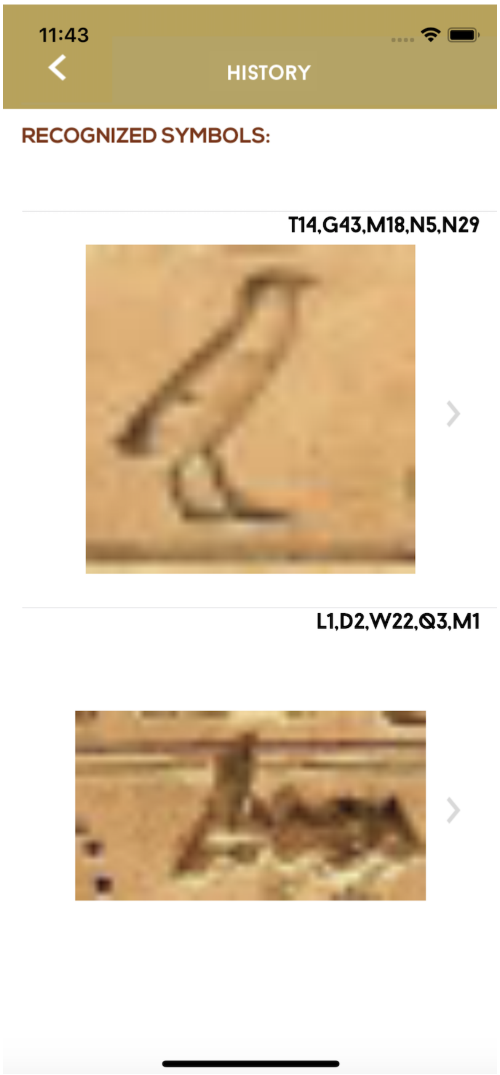
GARDINER'S CODE:
G14

DESCRIPTION:
Vulture

MEANING:
Mother

USEFUL LINKS:
Info about symbol:
http://www.egyptianhieroglyphs.net/gardiners-sign-list/birds/

(b) Symbol Dictionary



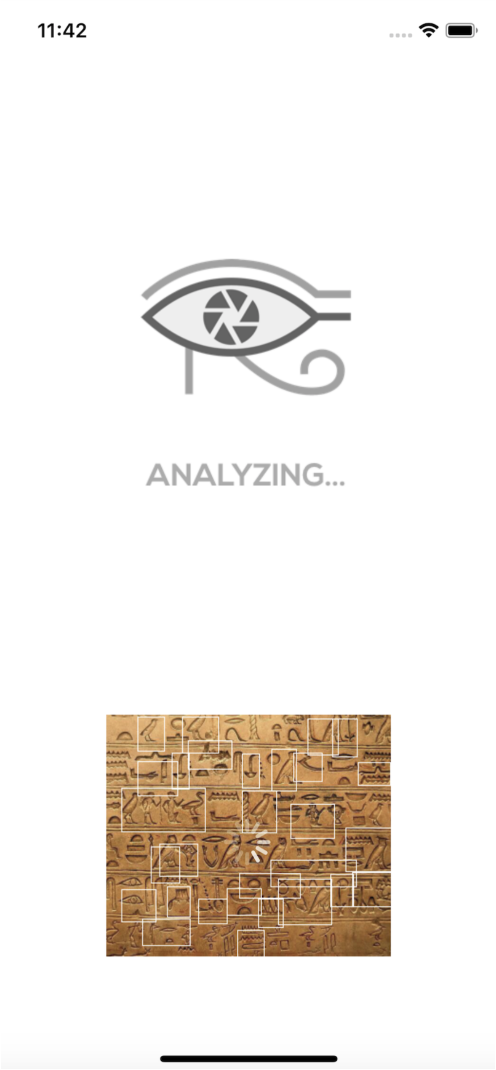
11:43

RECOGNIZED SYMBOLS:

T14,G43,M18,N5,N29

L1,D2,W22,Q3,M1

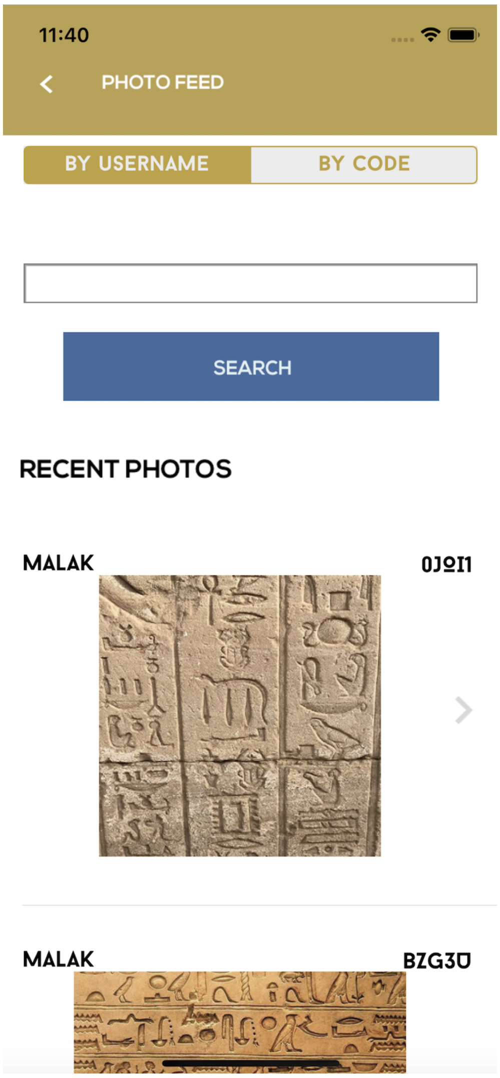
(a) Symbols recognized in photo



11:42

ANALYZING...

(b) Analyzing Progress Screen



11:40

PHOTO FEED

BY USERNAME | BY CODE

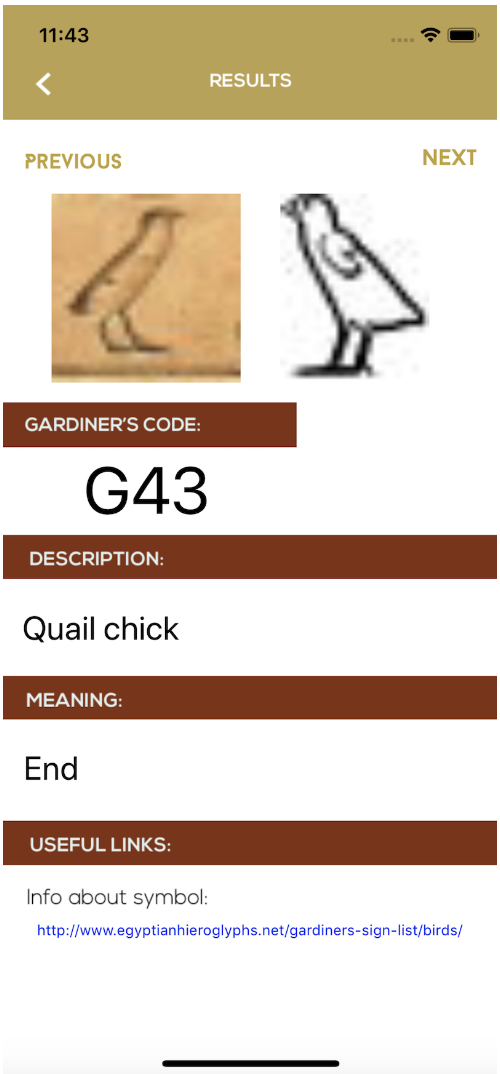
SEARCH

RECENT PHOTOS

MALAK 0J오I1

MALAK BZG3Ʊ

(a) Social Feed Screen



11:43

RESULTS

PREVIOUS NEXT

GARDINER'S CODE:

G43

DESCRIPTION:

Quail chick

MEANING:

End

USEFUL LINKS:

Info about symbol:

http://www.egyptianhieroglyphs.net/gardiners-sign-list/birds/

(b) Symbol Prediction Screen Result