Translating Ancient Egyptian Hieroglyphs into English

Reem Hussin*, Mahmoud Essa†, Mohanned Mahmoud‡, Abdelrahman Ahmed§, Mohamed Nehad¶
Faculty of Information Technology and Computer Science
December 2025

***Abstract*—Ancient Egyptian hieroglyphs constitute a complex pictorial writing system whose interpretation requires specialized expertise and extensive manual effort. Automating the translation of hieroglyphic inscriptions into modern languages remains challenging due to visual complexity, data scarcity, and the lack of large-scale parallel corpora. In this paper, we present an end-to-end framework for translating ancient Egyptian hieroglyphs from images directly into English by integrating Computer Vision and Natural Language Processing techniques.**

The proposed system employs a YOLO-based object detection model trained on synthetically generated inscription images to localize hieroglyphic symbols. Detected symbols are normalized using Gardiner codes, which serve as a stable intermediate representation bridging visual recognition and linguistic translation. A transformer-based sequence-to-sequence model based on BART is then fine-tuned to translate sequences of Gardiner codes into fluent English sentences. Synthetic data generation is leveraged to overcome annotation limitations and enable scalable detector training across approximately 310 hieroglyph classes.

Experimental results demonstrate that the system produces coherent and semantically meaningful translations despite limited training data, achieving a BLEU score of 62.17 and a ChrF score of 72.37 on the translation task. Qualitative end-to-end examples further illustrate the effectiveness of the proposed pipeline. This work is presented as a project-oriented report that documents the practical integration of modern deep learning techniques for hieroglyphic translation and provides a foundation for future research in digital Egyptology.

***Index Terms*—Hieroglyphic Translation, Computer Vision, YOLO, Sequence-to-Sequence Models, Low-Resource NLP**

I. Introduction

Ancient Egyptian hieroglyphs represent one of the earliest and most complex writing systems in human history. Used for more than three millennia, hieroglyphs were carved or painted on monuments, temple walls, tombs, and papyri to record religious texts, historical events, and daily life. Despite their cultural and historical significance, interpreting hieroglyphic inscriptions remains a difficult and time-consuming task that requires extensive expertise in Egyptology.

Automatic translation of hieroglyphs into modern languages presents several challenges. First, hieroglyphic writing is inherently pictorial, requiring reliable visual recognition before any linguistic processing can occur. Second, the language is considered extremely low-resource from a computational perspective, as large-scale parallel corpora are scarce. Third, a single hieroglyphic inscription may contain hundreds of distinct symbols arranged in complex spatial layouts, making end-to-end automation particularly challenging.

Most prior research has focused either on the visual recognition of individual hieroglyphic symbols using image processing or deep learning techniques, or on the linguistic translation of symbolic representations using Natural Language Processing (NLP) models. However, relatively few works address the full pipeline that starts from raw images of inscriptions and ends with fluent English translations. Recent studies have shown that combining Computer Vision (CV) and NLP techniques can significantly improve performance in low-resource and ancient language settings [1], [2].

In this paper, we present an end-to-end framework for translating ancient Egyptian hieroglyphs from images directly into English text. Our approach integrates object detection, symbolic normalization, and sequence-to-sequence translation into a unified pipeline. Hieroglyphic symbols are first detected from input images using a YOLO-based object detection model trained on a synthetically generated dataset. Detected symbols are then mapped to standardized Gardiner codes, which serve as a compact and linguistically meaningful intermediate representation. Finally, a BART-based sequence-to-sequence model translates the resulting symbol sequence into fluent English.

The proposed system is designed as a project-oriented report that documents the full development pipeline, experimental setup, and observed results. Rather than proposing a new linguistic theory, this work aims to demonstrate how modern deep learning techniques can be practically combined to address the challenges of hieroglyphic translation in a data-scarce environment.

The main contributions of this paper can be summarized as follows:

- An end-to-end Computer Vision and NLP pipeline for translating hieroglyphic images into English.
- A YOLO-based hieroglyph detection model trained using a large synthetically generated dataset covering approximately 310 symbol classes.
- The use of Gardiner codes as a normalization layer bridging visual recognition and language translation.
- A BART-based sequence-to-sequence translation model trained to map hieroglyphic symbol sequences to English sentences.
- A comprehensive experimental evaluation including qualitative translation examples and quantitative metrics such as BLEU, ChrF, exact match, and token accuracy.

The remainder of this paper is organized as follows. Section II reviews related work in hieroglyph recognition and translation. Section III describes the datasets and resources used in this project. Section IV details the proposed methodology and system architecture. Section V presents the experimental setup and results. Section VI discusses the findings and limitations, and Section VII concludes the paper with directions for future work.

II. Related Work

Research on the automatic processing of ancient Egyptian hieroglyphs spans several domains, including image processing, object detection, optical character recognition (OCR), and natural language translation. Due to the pictorial nature of hieroglyphs and the scarcity of large-scale annotated corpora, most existing approaches address only partial components of the overall translation pipeline.

A. Hieroglyph Recognition and Image-Based Approaches

Early studies primarily focused on recognizing individual hieroglyphic symbols using traditional image processing techniques. Elnabawy *et al.* [3] proposed an OCR-based framework that segments hieroglyphic symbols from images and maps them to English using Gardiner's sign list. While effective in controlled settings, this approach relied heavily on handcrafted features and rule-based segmentation, limiting its robustness and scalability.

Other works explored visual descriptor-based recognition. Franken and Van Gemert employed Histogram of Oriented Gradients (HOG) and spatial matching techniques to classify hieroglyphs, achieving moderate accuracy when symbols were manually segmented. However, performance degraded significantly for automatically segmented glyphs, highlighting the difficulty of reliable symbol extraction in real-world inscriptions.

With the advancement of deep learning, convolutional neural networks (CNNs) have been increasingly adopted for hieroglyph recognition. Barucci *et al.* demonstrated that CNN architectures such as ResNet, Inception, and a custom GlyphNet model can achieve high accuracy on isolated hieroglyph classification tasks. Nevertheless, these approaches typically focus on single-symbol recognition rather than full inscription analysis.

B. Object Detection for Hieroglyph Localization

To address the limitations of isolated symbol classification, recent studies have introduced object detection frameworks capable of locating multiple hieroglyphs within a single image. Nasser *et al.* [4] employed contour-based segmentation and Detectron2 for glyph detection prior to translation. Similarly, Sobhy *et al.* [1] proposed an end-to-end system that integrates glyph detection and translation, reporting a BLEU score of 42.2.

YOLO-based object detectors have gained particular attention due to their efficiency and accuracy. Bin Umer [5] conducted a comparative study of SSD, YOLOv5, YOLOv8, and Fast R-CNN for hieroglyph detection, demonstrating that YOLO-based models outperform other architectures in terms of mean Average Precision (mAP) and inference speed. These findings align with the broader success of YOLO architectures in general object detection tasks, as introduced by Redmon *et al.* [6] and further improved by Bochkovskiy *et al.* [7].

C. Hieroglyph Translation and NLP Models

Beyond visual recognition, translating hieroglyphic symbols into natural language introduces additional linguistic challenges. Rule-based translation systems that rely on deterministic mappings from Gardiner codes to English glosses offer high precision but lack flexibility and fluency.

Recent work has explored neural machine translation for ancient languages. De Cao *et al.* [2] proposed a transformer-based model adapted from the M2M-100 multilingual translation framework, demonstrating the effectiveness of transfer learning for hieroglyph translation in low-resource settings. However, their approach focused primarily on textual hieroglyph representations rather than direct image-to-text translation.

Sequence-to-sequence transformer models such as BART [8] have shown strong performance in text generation and translation tasks, particularly when fine-tuned on domain-specific datasets. Large multilingual models such as M2M-100 [9] further highlight the potential of pre-trained transformers for handling under-resourced and ancient languages.

D. Positioning of This Work

In contrast to prior studies, this project presents a unified end-to-end framework that translates hieroglyphic images directly into English text. By integrating YOLO-based object detection, Gardiner code normalization, and a BART-based sequence-to-sequence translation model, the proposed system bridges the gap between visual recognition and linguistic interpretation. The work is positioned as a comprehensive project report that demonstrates the practical integration of modern Computer Vision and NLP techniques for hieroglyphic translation in a low-resource setting.

III. Dataset and Resources

This section describes the datasets and external resources used to develop and evaluate the proposed hieroglyph-to-English translation system. Due to the low-resource nature of ancient Egyptian hieroglyphs, multiple datasets were combined, augmented, and synthesized to support both visual recognition and language translation tasks.

A. Hieroglyph Image Dataset

The core visual resource used in this work is a hieroglyph image dataset consisting of approximately 310 distinct symbol classes. Each class corresponds to a unique hieroglyphic sign mapped to its standardized Gardiner code. The dataset was sourced from an open repository of Egyptian hieroglyphic symbols and curated to ensure consistency in labeling and visual quality.

The images represent isolated hieroglyphic symbols extracted from inscriptions and manuscripts. These symbol crops serve as the foundational visual vocabulary for the system and are used for two primary purposes: (1) training baseline classification models and (2) generating synthetic inscription images for object detection.

Table I summarizes the key characteristics of the hieroglyph image dataset.

TABLE I
SUMMARY OF THE HIEROGLYPH IMAGE DATASET

Attribute	Value
Number of symbol classes	∼310
Image type	Isolated glyph crops
Labeling scheme	Gardiner codes
Primary usage	Detection and classification

B. Synthetic YOLO Detection Dataset

To overcome the lack of large-scale annotated inscription images, a synthetic dataset was generated to train the object detection model. Using the isolated glyph images, synthetic inscription-like images were programmatically constructed by placing multiple hieroglyph symbols on randomly generated stone-texture backgrounds.

Each synthetic image contains a variable number of hieroglyphs, randomly scaled and positioned to simulate real-world inscriptions. Corresponding bounding box annotations were automatically generated in YOLO format, enabling supervised training without manual labeling.

The synthetic dataset contains approximately 5,000 images and was split into training and validation subsets. This approach allows the object detector to learn robust localization patterns while significantly reducing annotation effort.

Figure 1 illustrates an example of a synthetically generated image along with its corresponding YOLO annotations.

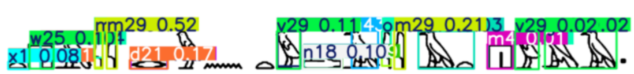

Fig. 1. Example of a synthetically generated hieroglyph inscription image with YOLO bounding box annotations.

C. Parallel Hieroglyph–English Translation Dataset

For the translation component, a parallel dataset was constructed linking hieroglyphic symbol sequences to English sentences. Each data instance consists of a sequence of Gardiner codes representing a hieroglyphic inscription and a corresponding English translation.

The dataset contains approximately 400 sentence pairs and was curated to ensure full alignment between the visual symbol classes and the translation vocabulary. All hieroglyphic symbols appearing in the translation dataset are included in the 310-class image dataset, ensuring consistency across the pipeline.

This dataset serves as the primary training and evaluation resource for the sequence-to-sequence translation model.

TABLE II
SUMMARY OF THE TRANSLATION DATASET

Attribute	Value
Number of sentence pairs	∼400
Source representation	Gardiner code sequences
Target language	English
Usage	Seq2Seq training and evaluation

D. Gardiner Code Knowledge Base

A knowledge base derived from Alan Gardiner's sign list was used to standardize hieroglyphic symbols across all system components. This resource provides a consistent mapping between visual symbol classes, Gardiner codes, and their linguistic descriptions.

The knowledge base acts as a normalization layer between the Computer Vision and NLP modules. After detection, predicted symbol classes are converted into their corresponding Gardiner codes, which form the input sequence for the translation model. This intermediate representation reduces visual noise and enforces linguistic consistency.

E. Dataset Alignment and Consistency

A key design decision in this project was to maintain full alignment between the visual and textual datasets. All 310 hieroglyph classes used for detection are represented in the translation dataset, ensuring that any detected symbol can be processed by the translation model. This alignment is essential for enabling end-to-end image-to-text translation without requiring manual intervention.

By combining curated visual data, synthetic augmentation, and parallel textual resources, the proposed system establishes a coherent and scalable dataset foundation for hieroglyphic translation.

IV. METHODOLOGY

This section describes the proposed end-to-end methodology for translating ancient Egyptian hieroglyphs from images into English text. The system is designed as a multi-stage pipeline that integrates Computer Vision and Natural Language Processing components, with Gardiner codes serving as an intermediate symbolic representation.

A. System Overview

The overall architecture of the proposed system is illustrated in Fig. 2. The pipeline begins with a raw image of a hieroglyphic inscription and produces a fluent English translation as output. The system consists of four main stages: (1) hieroglyph detection, (2) symbol normalization, (3) sequence construction, and (4) sequence-to-sequence translation.

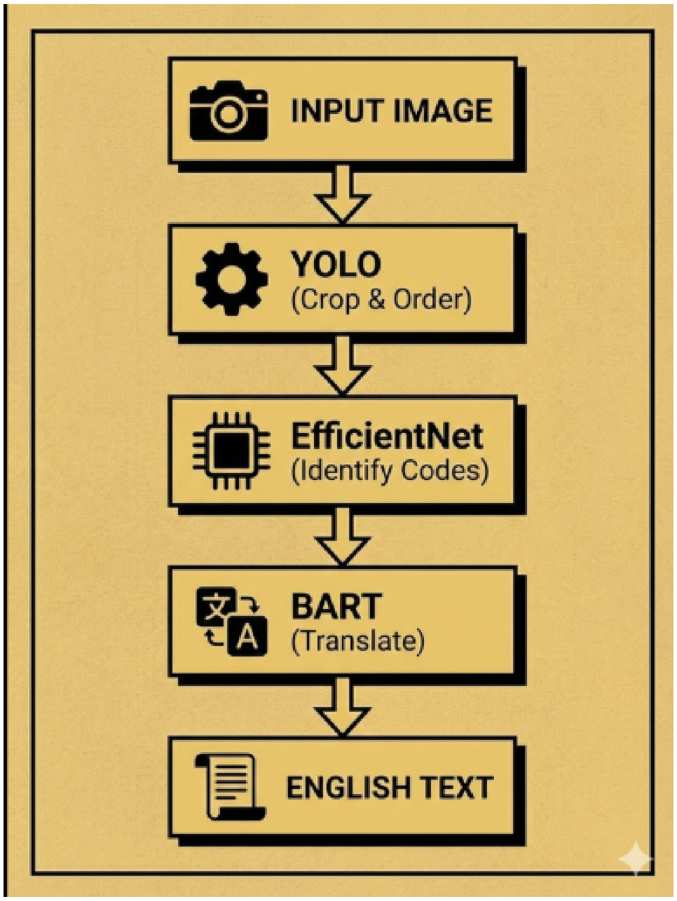

Fig. 2. Architecture of the proposed hieroglyph-to-English translation system.

B. Hieroglyph Detection Using YOLO

To localize hieroglyphic symbols within inscription images, an object detection model based on the YOLO architecture was employed. YOLO was selected due to its efficiency, high detection accuracy, and suitability for real-time applications.

The detection model was trained on a synthetically generated dataset constructed from isolated hieroglyph symbol images. Each synthetic image contains multiple hieroglyphs placed on stone-texture backgrounds, simulating real-world inscriptions. Bounding box annotations were automatically generated in YOLO format, eliminating the need for manual labeling.

The detector outputs bounding boxes along with predicted class labels corresponding to hieroglyph symbol classes. Each class is uniquely associated with a Gardiner code.

TABLE III
YOLO DETECTION MODEL CONFIGURATION

Parameter	Value
Model architecture	YOLOv8n
Input image size	960×960
Training epochs	50
Batch size	8
Confidence threshold	0.25
Number of classes	$\sim$310

C. Symbol Normalization and Gardiner Code Mapping

Following detection, each predicted bounding box is associated with a hieroglyph class label. These labels are mapped to standardized Gardiner codes using a predefined knowledge base derived from Alan Gardiner's sign list.

This normalization step serves as a critical bridge between the visual and linguistic components of the system. By converting raw visual predictions into Gardiner codes, the system enforces a consistent symbolic representation and reduces the impact of visual noise on the translation stage.

D. Symbol Ordering and Sequence Construction

Detected hieroglyphs are spatially ordered to form a symbolic sequence suitable for language translation. Bounding boxes are sorted based on their spatial coordinates using a simple geometric heuristic that prioritizes horizontal position followed by vertical position.

The resulting ordered list of Gardiner codes forms the input sequence for the translation model. This approach ensures reproducibility and avoids introducing language-specific reading rules, allowing the system to generalize across diverse inscription layouts.

E. Sequence-to-Sequence Translation Using BART

The translation stage employs a sequence-to-sequence transformer model based on the BART architecture. The model is fine-tuned to map sequences of Gardiner codes to English sentences.

BART was selected due to its strong performance in generation and translation tasks, particularly in low-resource settings. During training, Gardiner code sequences are tokenized and provided as input to the encoder, while the decoder generates the corresponding English translation.

Beam search decoding is applied during inference to improve translation quality and stability.

TABLE IV
BART TRANSLATION MODEL CONFIGURATION

Parameter	Value
Base model	BART
Training epochs	20
Batch size	16
Learning rate	3×10^{-5}
Decoding strategy	Beam search
Number of beams	4
Maximum output length	64 tokens

F. End-to-End Inference Pipeline

During inference, a user provides an image containing a hieroglyphic inscription. The image is first processed by the YOLO detector to localize and classify hieroglyphs. Detected symbols are normalized to Gardiner codes, spatially ordered, and assembled into a sequence. This sequence is then passed to the BART model, which generates the final English translation.

Figure 3 shows an example of the full pipeline, including detection results and the corresponding translation output.

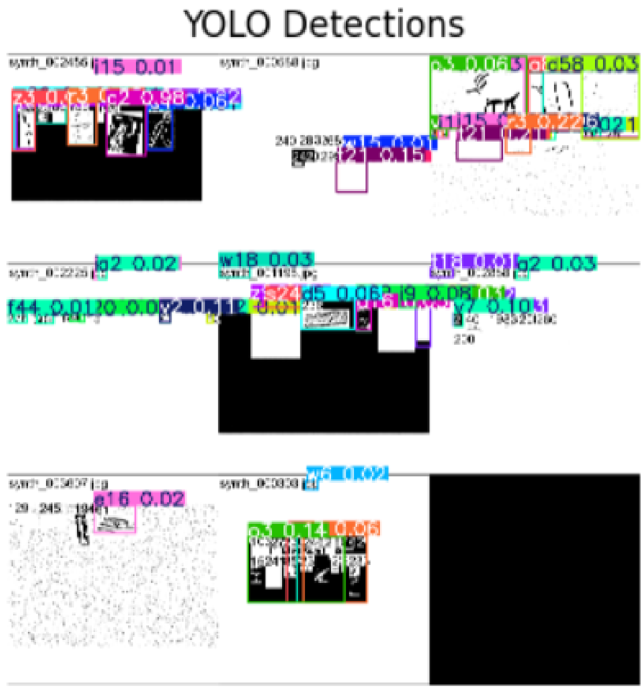

Fig. 3. Example of end-to-end translation from a hieroglyphic image to English text.

Codes: ['o3', 'e16', 'r3', 'w18', 'c2', 'a4', 'i15', 'n35', 't21', 's24', 'r3', 'd39', 'g9', 't21', 'v2', 's24', 'd5', 'a16', 'v29', 'i9', 'v7', 'a24', 'w18', 'o3', 'r3'] English: He drinks water carefully.

The modular design of the proposed methodology allows each component to be independently improved or replaced, while maintaining the overall structure of the translation pipeline.

V. Experimental Setup and Results

This section presents the experimental configuration used to evaluate the proposed hieroglyph-to-English translation system and discusses both quantitative and qualitative results obtained from the end-to-end pipeline.

A. Experimental Setup

All experiments were conducted in a GPU-enabled cloud environment. The Computer Vision and Natural Language Processing components were trained independently and then integrated into a unified inference pipeline.

The hieroglyph detection model was trained using the synthetically generated dataset described in Section III, while the sequence-to-sequence translation model was trained on the parallel Gardiner–English corpus. Evaluation was performed on both synthetic inputs and real hieroglyphic images to assess the robustness of the overall system.

TABLE V
Experimental Environment and Setup

Component	Configuration
Frameworks	PyTorch, Ultralytics YOLO, HuggingFace Transformers
Hardware	GPU-enabled cloud environment
Platform	Google Colab
Evaluation data	Synthetic and real hieroglyphic inputs

B. Translation Model Quantitative Evaluation

The translation component of the system was evaluated using standard machine translation metrics, including BLEU, ChrF, exact match accuracy, and token-level accuracy. These metrics provide complementary insights into translation quality, fluency, and semantic correctness.

TABLE VI
Translation Model Evaluation Results

Metric	Score
BLEU	62.17
ChrF	72.37
Exact Match	0.34
Token Accuracy	0.78

The obtained BLEU and ChrF scores indicate that the model produces fluent and semantically meaningful English translations despite the limited size of the training dataset. The relatively lower exact match score reflects the inherent variability of natural language generation, where multiple valid translations may exist for a given hieroglyphic sequence.

C. End-to-End Qualitative Results

To demonstrate the effectiveness of the full pipeline, qualitative end-to-end translation results are presented in Fig. 4 and Fig. 5. Each example shows the Gardiner code sequence extracted after the vision and normalization stages, followed by the generated English translation.

```
Codes: ['o3', 'e16', 'r3', 'w18', 'c2', 'a4', 'i15', 'n35', 't21', 's24', 'r3', 'd39', 'g9', 't21', 'v2', 's24', 'd5', 'a16', 'v29', 'i9', 'v7', 'a24', 'w18', 'o3', 'r3']
English: He drinks water carefully.
```

Fig. 4. End-to-end translation example showing the extracted Gardiner code sequence and the generated English sentence: “Wash the map in silence.”

```
Codes: ['r3', 'w18', 'n35', 'v4', 'o2', 'w18', 'r3', 'w18', 'r3', 'w18', 'r3', 'w18', 'i14', 'e23', 'n39', 'r3', 'w18', 'r3', 'm42', 'r3', 'o2', 'w18', 'o2']
English: Wash the map in silence.
```

Fig. 5. End-to-end translation example showing the extracted Gardiner code sequence and the generated English sentence: “He drinks water carefully.”

These examples demonstrate that the proposed system is capable of producing coherent and contextually appropriate English translations from hieroglyphic inputs. Even when the extracted symbol sequences contain repetitions or visually similar signs, the sequence-to-sequence translation model shows robustness in generating meaningful outputs.

Overall, the experimental results confirm that combining object detection, symbolic normalization, and transformer-based translation enables effective end-to-end hieroglyphic translation in a low-resource setting.

VI. Discussion and Limitations

The experimental results demonstrate that integrating Computer Vision and Natural Language Processing techniques enables effective end-to-end translation of ancient Egyptian hieroglyphs into English. By decomposing the task into symbol detection, symbolic normalization, and sequence-to-sequence translation, the proposed system successfully addresses several challenges inherent to hieroglyphic writing, including visual complexity and data scarcity.

One of the key strengths of the proposed approach lies in the use of synthetic data for training the object detection model. Synthetic data generation significantly reduces the need for manual annotation while allowing the detector to learn robust spatial and visual patterns. This strategy proved particularly effective given the large number of hieroglyph classes and the limited availability of labeled real-world inscription images.

The use of Gardiner codes as an intermediate representation also plays a crucial role in system stability. By mapping visual predictions to standardized symbolic identifiers, the system isolates visual noise from the translation model and enforces linguistic consistency. This design choice simplifies the translation task and improves generalization, especially in low-resource settings.

Despite these strengths, the proposed system has several limitations. First, the spatial ordering of detected symbols is based on a simple geometric heuristic rather than explicit linguistic rules. While this approach provides consistency and reproducibility, it does not fully capture the complex reading conventions of ancient Egyptian inscriptions, which may vary depending on layout and context.

Second, the translation model is trained on a relatively small parallel dataset. Although the achieved BLEU and ChrF scores indicate strong performance, the limited size of the dataset restricts the system's ability to handle rare symbol combinations and complex grammatical structures. Larger and more diverse corpora would likely improve translation accuracy and fluency.

Third, errors introduced during the detection stage can propagate through the pipeline and affect translation quality. Misclassified or missed symbols may lead to incomplete or incorrect Gardiner code sequences, which in turn impact the generated English output. While the translation model exhibits some robustness to noise, it cannot fully compensate for significant detection errors.

Finally, the system focuses on translating hieroglyphic sequences into modern English without explicitly modeling deeper linguistic phenomena such as morphology, syntax, or semantic roles. As a result, some translations may lack nuance or stylistic precision, particularly for longer or more complex inscriptions.

Overall, the proposed system should be viewed as a practical and extensible framework rather than a definitive solution to hieroglyphic translation. The limitations identified in this section highlight important directions for future research and system improvement.

VII. Conclusion and Future Work

This paper presented a comprehensive end-to-end framework for translating ancient Egyptian hieroglyphs from images into English text by integrating Computer Vision and Natural Language Processing techniques. The proposed system combines object detection for hieroglyph localization, symbolic normalization using Gardiner codes, and a transformer-based sequence-to-sequence translation model to address the challenges of hieroglyphic translation in a low-resource setting.

Through the use of synthetic data generation, the system effectively mitigates the lack of large-scale annotated hieroglyphic image datasets, enabling the training of a robust detection model across approximately 310 symbol classes. The adoption of Gardiner codes as an intermediate representation provides a stable and linguistically meaningful interface between visual recognition and language translation. Experimental results demonstrate that the proposed approach can generate coherent and semantically meaningful English translations, achieving competitive BLEU and ChrF scores despite limited training data.

While the system is presented as a project-oriented report rather than a production-ready solution, it demonstrates the feasibility of combining modern deep learning techniques to support the interpretation of ancient writing systems. The modular design of the pipeline allows individual components to be independently improved or replaced without altering the overall architecture.

Future work may focus on several directions. Expanding the parallel hieroglyph–English corpus would improve translation coverage and linguistic accuracy. Incorporating more advanced reading-order modeling and contextual layout analysis could better capture the structural conventions of hieroglyphic inscriptions. Additionally, integrating linguistic knowledge such as grammatical rules or semantic constraints may enhance translation quality for longer and more complex texts. Finally, extending the system to support additional target languages or interactive educational applications represents a promising avenue for further development.

In conclusion, this work demonstrates that an end-to-end image-to-text translation pipeline for ancient Egyptian hieroglyphs is both achievable and effective using contemporary Computer Vision and NLP methodologies, offering a practical foundation for future research in digital Egyptology.

References

[1] A. Sobhy, M. Helmy, M. Khalil, S. Elmasry, and N. Negied, "An ai-based automatic translator for ancient hieroglyphic language—from scanned images to english text," *IEEE Access*, vol. 11, 2023.

[2] M. De Cao, N. De Cao, A. Colonna, and A. Lenci, "Deep learning meets egyptology: A hieroglyphic transformer for translating ancient egyptian," in *Proceedings of the Workshop on Machine Learning for Ancient Languages (ML4AL)*, 2024.

[3] R. Elnabawy, R. Elias, M. A.-M. Salem, and S. Abdennadher, "Extending gardiner's code for hieroglyphic recognition and english mapping," *Multimedia Tools and Applications*, 2020.

[4] A. Nasser, M. Mohamed, A. Sherif, and B. Mahmoud, "Hieroglyphtranslator: Automatic recognition and translation of egyptian hieroglyphs to english," *arXiv preprint arXiv:2512.03817*, 2025.

[5] W. Bin Umer, "Translating egyptian hieroglyphs using deep learning," Master's thesis, National College of Ireland, 2024.
[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[7] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
[8] M. Lewis, Y. Liu, N. Goyal *et al.*, "Bart: Denoising sequence-to-sequence pre-training for natural language generation," *Proceedings of ACL*, 2020.
[9] A. Fan, S. Bhosale, H. Schwenk *et al.*, "Beyond english-centric multilingual machine translation," *Journal of Machine Learning Research*, 2021.