# Translating Egyptian Hieroglyphs Using Deep Learning

MSc Research Project
Data Analyst

Waleed Bin Umer
Student ID: X23187956

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

**Student Name:** ....... ................................ Waleed Bin Umer ............................................

**Student ID:** ........................................... X23187956 .........................................................

**Programme:** ......... Data Analyst ........................................ **Year:** .........2024...............

**Module:** .................................. MSc Research Project ...............................................

**Supervisor:** ..................................... Vladimir Milosavljevic ...........................................

**Submission Due Date:** ..............................................12-12-2024..................................................

**Project Title:** ......... Translating Egyptian Hieroglyphs Using Deep Learning.......................................

**Word Count:** ......... 8420................. **Page Count**.............................24..................

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** ..................................... Waleed Bin Umer ...........................................

**Date:** ................................................12-12-2024...........................................

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Table of Contents

# Translating Egyptian Hieroglyphs Using Deep Learning

Waleed Bin Umer
X23187956

**Abstract**

Egyptian hieroglyphs that are known for its diverseness has captured the attention of humans for centuries. People working for research or visiting Egypt as tourists are interested equally in knowing more about the insights it holds which is why the tourists are ready to pay high amounts to local guides which helps in translating and understanding. The deep learning models that are used are SSD, YOLOv5, YOLOv8 and Fast R-CNN which identify challenges in Egyptian hieroglyphs and translate them. All these images are curated and pre-processed using different techniques like augmentation, normalization and resizing, all of this is possible via training. All these models are evaluated based on their expertise to function in different systems but they do have the limitations within them. The mean average precision (mAP) of every model varies from backgrounds and in complex situations which determines their capability to work in different settings as mAP of YOLOv8 was 0.975, that outperformed among all models and YOLOv5 also performed well having mAP of 0.970. The Fast R-CNN had mAP of 0.937 and SSD had lowest mAP that is 67.54. This study also explores the technique of optimization within the system for tourists to have seamless reliance on the guides that are local by focusing on the cultural heritage of the Egyptian hierarchy

## 1 Introduction

### 1.1 Background

The hieroglyphs of the Egyptian are well known for the way they are written. The symbols in them not only represent their culture but it also holds history and shows their religion side. Tourists visiting Egypt hire the local guides that demands high money as 400 to 500 dollars for helping in translating it for them. They take high pay but there is no surety that it is properly being translated. The hurdle is that this has been going on for years from as long has world has come into existence but the proper solution has yet not achieved and people had no focus on this side. In recent years different deep learning and the computer vision techniques are used in a way by which translation could be easily understood. Techniques like image recognition by using different deep learning models are done to understand modern languages but translating hieroglyphs was still a task

## 1.2 Motivation

Understanding hieroglyph is not very easy, so people that visit Egypt have no understanding towards it for which they have to hire local guides that are costly and they help them to understand it, which gets more complex because most of the time the language becomes the barrier. The main motive of this study is to contribute both the culture and technology field too. As the hieroglyphs contain symbols that hold history and to translate them was a task but it became possible by this study as all the models performed well and gave results accordingly. All four models such as SSD, YOLOv5, YOLOv8 and Fast R-CNN have shown their strength and limitations. Still key metrics analysis showed each model has good mean Average Precision and proper translation is done so that the culture could be preserved for years and the gap could be fulfilled not only for tourists but for researchers too.

## 1.3 Research Question

How can successful real-time translation of hieroglyphs be enhanced by using Deep learning algorithms so that the tourists can have flawless experience on mobile phones while visiting Egypt?

This research question focus is on applying the deep learning algorithms for various models so that proper identification of the Egyptian hieroglyphs can be done. In addition, the comparison of models is also achieved so that accurate answers could be given to the tourists.

## 1.4 Research Objectives

Below are the objectives to address the above research question

- The first objective is the understanding of the related deep learning techniques which are used for the recognition of ancient scripts and looking into the implementation of different objection detection models like SSD, YOLOv8, YOLOv5 and Fast R-CNN. Examination of these models helps to identify the best suitable model for the mobile use while keeping a main focus on SSD-mobilenet V2 because its framework is lightweighted and shows high processing speed.
- The second objective is looking into the techniques like data augmentation and the processes of training that covers the gap between old history and technology. The deep learning models helps to make sure that efficient performance is done by models.
- The third objective is analyzing mAP, performance speed, the precision, recall value and F1 score value of every model had helped to identify the efficiency of Deep learning algorithm by analyzing and to know which one has more better results.

# 2 Related Work

## 2.1 Introduction

New technologies in the field of AI or related field like ML and DL has broaden in every field, many machine learning and deep learning techniques are providing tools in every field. This literature review investigates into the key findings of different deep learning techniques for the identification of Egyptian hieroglyph and this review also provide the related work of DL algorithms in the ancient scripts. Overall, this review focuses on the methodology results limitation of the different recent studies.

## 2.2 Key Findings of Related Algorithm in Detection Of Egyptian Hieroglyphs

In a study of Andera Barucci et.al, 2021 used CNN technique for the classification of Egyptian hieroglyph. CNN architecture is essential for identifying Egyptian hieroglyph. Image processing and Image augmentation techniques like scaling and rotating were applied to understand the data set stability which also helps in training of the model. Three CNN framework ResNet-50, Inception-V3 and Xception were used for the training and classification of the images. To modify the already existing network Glyph net CNN model was developed. Basically, Glyph net was developed to handle the complexity of ancient symbols. Then the comparison between other CNN models was done with Glyph net by using evaluation metrices which shows accuracy of 92.30%, precision of 90.50%, F1 score of 91.20%, and recall of 91.80%. where it was concluded that Glyphnet has given good performance. Overall, this research shows that transfer learning techniques or hybrid model helps to achieve good accuracy.

Many complex neural networks are used for executing these tasks but for implementing these tasks with deep networks require large datasets. p. Lion et.al,2024 in his study has discussed localization of hieroglyphs by using techniques of unsupervised learning where it has helped in differentiating the hieroglyphs for proper processing without annotated data. The Hough transform is used that helped in detecting lines and shapes in the images. The shapes including the edges are detected by using the canny edge algorithm.

Furthermore, for classifying and detecting of the hieroglyphs, the Fast R-CNN is a supervised model but, in this research, it is used as an unsupervised conveyor by using the outputs of edge and shape detection which helps to identify the localized features of hieroglyphs. This paper has highlighted the limitations of using unsupervised method like identifying the complex symbols the R-CNN model face difficulty with or without labelled data.

Another study of Reham Elnabawy et.al, 2021 proposes an algorithm which segment and identifies the ancient Egyptian hieroglyphs with the help of Gardiner’s list which helps in mapping the hieroglyphic symbols to translate in English. This study has its focus mainly on identifying and translation of hieroglyphs as they are important for educational purposes too. The Gardiner's list with the model of machine learning is used for identification of symbols. Technique like NLP has also been used for translating them into English language individually. Limitation of this study indicates the complexity of translating and identifying the hieroglyph. Overall, this approach is concluded as significant by indicating that NLP techniques is essential for translating hieroglyph in to a meaningful text. This mainly has its focus on writing where it is discussed that preserving the knowledge into written form has been a better way which is why doing perfect mapping from the scripts like this into the modern language should be focused in an automatic way so that the output could be got.

Whereas Morris Franken et.al,2013 does introduces an approach in a study which automatically identifies the Egyptian hieroglyphs through images and retrieve image as textual representation. Each image of the hieroglyph is understood and recognized by evaluating the five visual descriptors which are shape, texture, color, edge and spatial layout these descriptors help to convert hieroglyph in to textual representation but before that the hieroglyph is placed and segmented. The results include Average classification Rate for Manually cut-out Hieroglyphs and Average classification rate, which are 74% ± 1% and 53% ± 5%, suggests that this approach shows complexity in recognition of symbols especially for the hieroglyphs which are in common.

Furthermore, the study of Costanza cucci et.al,2024 in his work has used hyperspectral imaging (HIS) in the visible (Vis) and (NIR) known as Near infrared range integrated with convolutional neural network to recognize the hieroglyphs. The hyperspectral imaging in the Vis and NIR are used to recognize the ancient hieroglyphs that are difficult to read because in real time the surface area of the ancient hieroglyphs are damaged. The HIS technique enhances the readability of damaged hieroglyphs and the information retrieved from HIS is used to enhanced the images and fed to the CNN framework. The study highlights this method effectively detect the micro-level color and texture. The result of HIS and CNN is evaluated but it is not mentioned.

## 2.3 Mobile Application

On the other hand, Ragaa Moustafaa et.al,2022 in a paper proposes developing an application named as “Scriba” using flutter. It is mentioned that three CNN techniques like Efficient Net, Mobile Net, and Shuffle Net have been used and tested on two datasets of hieroglyphs with and without data augmentation both and results shows that Mobile Net and Efficient Net shows best accuracy which is vary from 99% to 100%. The application is developed to recognize hieroglyph and translate it into English language. CNN framework is used for symbol recognition and sequence-to-sequence model is used for the translation of the detected hieroglyphs. The results of this application suggest that it is significant for the identification of hieroglyphs.

## 2.4 Mobile Application Use of CNN And DL Based Approaches in Different Ancient Scripts

In a research Dr shikha verma et.al, 2022 has introduced a prediction and translation system that is of ancient scripts using AI. This research uses CNN networks along with language translation transformer models which helps to covert symbols into text. IM translator is used as initial layer in the model which is used for the recognition of symbols.IM translator is used to map the recognized symbols into the modern language. Findings shows the BLEU score of 8 % whereas the score of IM translator shows 0.650 and the (WER) shows 10% comparing with IM translator provides rate of 0.150, the analysis of N-gram represents increase of 3% to 4%. The result shows that AI approach is significant.

Furthermore, Amar Jindal et.al, 2024 proposed a hybrid deep learning model on the basis of OCR method to identify handwritten characters in ancient document image which is in two Indian scripts Devanagari and Maithili. The OCR method helps in extracting features with the help of Convolutional layers. The hybrid model be formed of on dense layer and a variety of connected hidden layers. Long short-term memory and Bidirectional long short-term memory alternation of recurrent neural network is also a applied in the portion of connected hidden layers. The evaluation of the OCR method has tested on two datasets which consist of Devanagari and Maithili ancient documents. The accuracy for the Devanagari is 96.97% and for Maithili the accuracy is 95.83%.the result shows that OCR method performs well.

In another study Sonika Rani Narang et.al,2021 also uses a deep learning approach for the recognition of Devanagari ancient character. CNN is used to extract the information from ancient Devanagari scripts. The researchers use model in different design options like number of layers, stride sizes, number of filters along with their kennel size Different experiments shows that CNN accomplish better accuracy for feature extraction. The model in this study is used as feature extractor and as a classifier as well. The model is trained for the recognition of 33 classes of characters and the results shows 98.72% of accuracy indicating that it effectively handles the different and unique pattern of the script.

Whereas the study of Alshahrani et.al, 2024 has applied a DL approach to convert a handwritten Arabic text into digital form. The deep learning technique combines with convolutional neural network and with Bidirectional long short-term memory and CTC, the model achieved an accuracy of 99.52% and two metrices WER and CER is used for evaluation the result of WER shows 1.640 and for CER it shows 0.485.

The methods mentioned above were applied by using the Character Error Rate and Word Error Rate metrics so that the performance could be compared giving better results for recognizing the handwritten text. The dataset was collected that had a vast range of text variation in Arabic and recognizing those Arabic words did lead to difficulties as understanding its script that follows the structure of letters being in a positioned way including the dots and hamzas is not easy.

The study of Akram Bennour et.al,2024 in this research has proposes a deep learning model for the writer identification of historical manuscripts with the help of data driven features. The purpose of the study is to recognize and learn the relevant features from the data and extinguish the need of manual feature engineering. The methodology of this work includes preprocessing of manuscripts which is done by image denoising using canny-edge detector. The extracted features were put through classification process with the help of deep learning model. Evaluation of this study shows 95.30% of accuracy rate, precision rate 94.80%, recall rate 95.10% and F1 score of 95.00%, results indicating that proposed DL framework is powerful to identify historical manuscripts.

Research of S. Umamaheswari et.al,2024 shows deep leaning approach to convert images of ancient Tamil scripts in to a modern Tamil text and the text can be translated in to any language with the help of google translate API. The process is accomplished by preprocessing of the images such as segmentation, noise reduction. The neural network is then applied once the preprocessing is completed that helps in recognizing the images and to take out the information from the image. Each image takes 1.8 second for processing. The model provides character recognition accuracy of 92.50%, word recognition accuracy shows 89.30%. This approach arises the complexity when translating the inscriptions especially those which are ancient and are mostly written on the stone or on metal being translated in the low resource languages. The webpage and web application are mentioned that can be used after the model training is done so that more audiences could be focused.

The study of Alam Ahmad Hidayat et.al, 2021 introduces a convolutional neural based classifier to recognize ancient Sundanese characters, acquired from digital collection of southeast Asian palm leaf manuscripts. The two preprocessing techniques applied for the training of images first technique includes geometric transformation and brightness adjustment, second technique include Otsu's transformation which is used to binarize the characters and uses the geometrical transformation for data augmentation. The framework is trained on the training and testing set. The evaluating result shows the second technique performs better than first technique that had the 97.74% accuracy.

Furthermore, Sakith Gunasekara et.al,2024 uses deep learning technique to develop mobile application that helps to recognize and translate ancient Brahmi characters in to modern language. This research uses data driven approach for the recognition and accurate prediction. (OpenCV) library is used for the preprocessing of image, helps to improve the quality of data, for the character recognition semantic segmentation techniques like TensorFlow and keras is used, and for the mobile app development flutter framework is used. The result shows an accuracy of 93.50%.

The research of Ruchika Malhotra et.al, 2023 has used a Deep learning technique to recognize historical Ethiopic texts. In study, the model exert end to end strategy which helps to enable sequential feature extraction and efficient recognition. In addition, the training data was increased by using data augmentation to address the issues of data shortage. The framework uses end-to- end strategy for the sequential feature extraction and for accurate recognition. The models attention mechanism combined with the connectionist temporal classification framework. The model contains 7 CNN and 2 recurrent neural network for feature extraction which is used to process sequential text. The evaluation of the model is tested on two test sets. Test Set I contain 6,150 samples. Where the character error rate is 17.95% and the TEST SET II contain 15,935 samples where the CER shows the 29.95% the test set II shows higher error rate indicating that there is need of improving the recognition historical text.

The paper of Deepali P. kadam et.al,2024 shows techniques of image analysis and AI in order to get correct information from the images of old ancient scripts documents. Adaptive gaussian thresholding is used for the clarity of image then (OCR) is applied through an advanced systems like Easy OCR and Tesseract helps images to transform into a readable text, assuring a good accuracy. NLP tools like happy transformer and hugging face is also applied for reducing the issues like blur text. The study shows an accuracy of 92.50%, the OCR rate (precision) shows 90.80% and OCR rate (Recall) of 91.20%, F1 score shows 91.00%

The study of A. Sarala et.al,2024 goes through detection of historic inscription images with the help of convolutional recurrent neural network. The model uses two powerful model the CNN and RNN where the Convolutional neural network is used for feature extraction and recurrent neural network helps to identify the sequences. The images in the dataset were manually annotated. The model training process includes fine tuning, hyperparameter optimization and connection temporal classification which works as an advanced technique, helps to handle the sequential data. After training the model is tested for evaluating the performance, shows 98.11% of accuracy.

The work of Alexandros Haliassos et.al,2020 uses new image processing and machine learning techniques for the classification and detection of ancient symbols in papyri. First color thresholding and contouring technique is used which helps in feature extraction of symbols. After feature extraction two classifiers are used: Support Vector Machine, and convolutional neural network. For the efficient performance of the system employ new technique, which is sliding window approach, helps to break images that contain relevant information. The system shows 81% of precision and recall of 74%.

The research of Mohammad Anwarul Islam et.al, 2023 has used the CNN model that has it focus on developing character recognition in the ancient documents. The limited size of data is used around 404 images. Due to limited size of the data a technique called data augmentation is applied by using python library known as image data generator. The training dataset was augmented three times and called as resampling 1,2,3. In a research customized CNN model is developed and compared to other machine learning techniques like support vector machine, k-nearest neighbour, Decision tree, Random Forest, XGBoost.

Further pre-trained models such as VGG16, MobileNet, and ResNet. The performance of the model was tested on the three samples of augmented dataset. The first sample of augmented dataset achieve accuracy of 88.67%, the second sample of augmented dataset shows 90.91% accuracy, the third sample shows 98.86%. The results shows that CNN performs better. To find out the efficiency of CNN model more, tested on MNIST dataset and the model achieved the accuracy on 99.03% on the MNIST.

# 3 Research Methodology

This section explains a proper workflow and techniques of deep learning that are used for the detection of Egyptian hieroglyphs. The methodology discusses about each step, it starts with the dataset preparation, then preprocessing of dataset was carried out. Where steps of EDA and annotation steps were performed to ensure the reliability of the data. Data augmentation is used to enhance the data, and to optimize learning the techniques of feature engineering is employed. The models SSD, YOLOv8, YOLOv5, and Fast R-CNN were used to train the processed data. The performance of the models was trained using the evaluation metrices includes mean average precision, precision, recall and F1. For the validation of prediction visualization were created, below the figure1 shows the overall methodology of this approach.

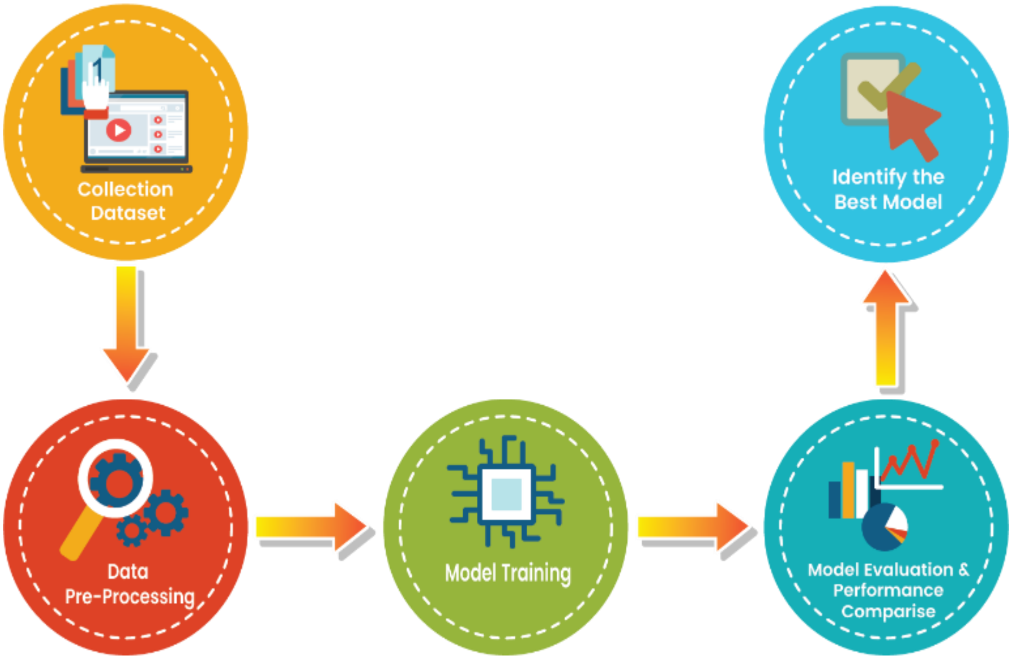



**Figure 1: Methodology Flow Diagram for Translating Egyptian Hieroglyphs Using Deep Learning.**

## 3.1 Dataset Description

The dataset is achieved from the Kaggle, and it is composed 8000 of Egyptian hieroglyphic images and the images are annotated for the object detection. It contains 95 unique classes of hieroglyphic images that were used in training phase and for the evaluation of implemented models. Every image is annotated through bounding box along with the class labels and are formatted in Pascal VOC standard that allows to identify the coordinates of bounding box such as (xmin, ymin, xmax, ymax). This annotation way makes training phase easier for the models. The analysis of dataset was conducted through the execution of EDA, shows that some hieroglyphs are repeated. For example, common symbols like Ankh and eye of Hours were represented explicitly but the quantity of other symbols was limited. This class imbalance caused challenge in training phase. Moreover, the analysis of bounding box ranges of different box sizes which cause difficulty in detection of smaller objects and to overcome this the techniques of data augmentation were employed which helps to enhance the quality of data. The dataset had complexity due to its having variety of images. As many images had symbols that were overlapped that caused the dataset to be diverse. This diversity led to complications but gave generalization that provided ease in preprocessing of the data. For effectively training of model, it is important to make sure that dataset is in proper way.

## 3.2 Data Pre-Processing

The steps in preprocessing have been carried out in a manner that there is proper reduction in the challenges that might have been raised during analysis of the data. Image resizing, normalization and reformatting of the bounding box was also applied for transformation of the raw images. The image resizing to the input dimension is the crucial step in every model. For the SSD model the resizing of the image required at 300x300 pixels and for YOLOv5 it requires 416x416 pixels,320X320 pixels are required for YOLOv8, and Fast R-CNN requires 800X800 pixels. The image resizing ensures the efficient training. The range of pixel values were checked between [0,1] in normalization. The normalization helps model to improve and allows to learn from the data easily by making sure that no pixel intensity is affected. The formatting of bounding box was carried out according to the requirements of the model. The coordinates from the corner were maintained for the SSD model and for the YOLO model coordinates from the center were maintained. Which ensures that all the annotations are reliable to go with the algorithm that has been used by each model. After that, data augmentation was carried out to balance the classes and for the improvement of model generalization. In this step the geometric transformation was performed which includes random rotations, random zooms and flipping(horizontal), for carrying different light conditions color adjustments was done. The dataset was divided into sets of training, validation sets having a ratio of 80:20. The division of the dataset assure that dataset can be trained validated on a distinct set of data, which helps to avoid it from overfitting, and provides reliability.

## 3.3 Data Visualization

Exploratory data analysis allows to uncover the patterns and extract important features from it that helps in preprocessing. The steps of EDA include class distribution, bounding box analysis and spatial distribution. In class distribution class imbalances were identified. The dominating classes of hieroglyph were visualized through bar graph. The head () and info () functions were used to explore the training, validation sets. The exploration of these sets shows training set contains 95 classes and validation set have 96 classes, in comparison of training set with validation set indicates that it contains 9 extra classes. Principal Component Analysis was used to carry out dimensionality analysis which shows variability of data. Dimensions of bounding box was analyzed through scatter plots which indicates hieroglyphs width and height within certain range. Figure 2 represents the scatter plot where the object detection has done at the center in a uniformly way and detecting the hieroglyph within the image space. Figure 3 highlights the bounding box to image ratio which represents the distribution of bounding box according to the image sizes, showing that higher concentration of bounding boxes is closer to the center of the images. To inspect the hieroglyphs spatial distribution Heatmaps were employed which helps to highlights the clustering patterns. Figure 4 shows spatial distribution that shows objects were dense.

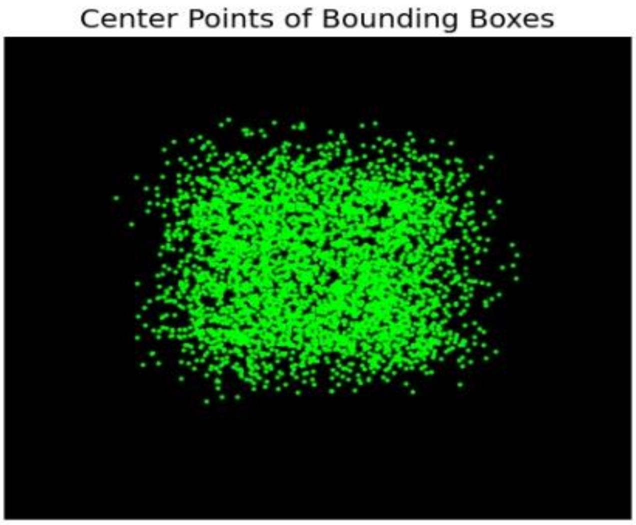



**Figure 2: A Scatter Plot**

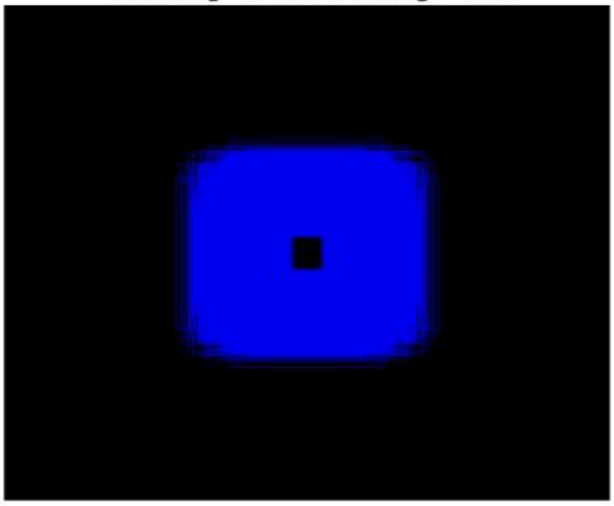



**Figure 3: Visualization of blue bounding boxes**

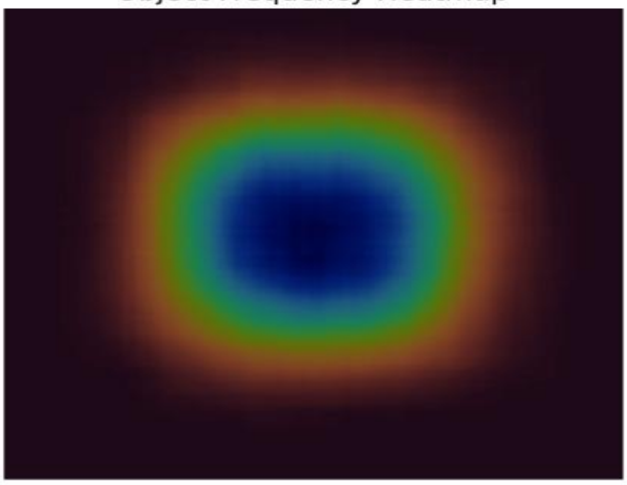



**Figure 4: A heatmap**

## 3.4 Model Training

The four models were trained where accuracy of four models and reliability of object detection models was checked. The dataset was divided into 80% and 20% for training and validation set. Where the training is used for optimizing weights. The validation set is used for to reduce overfitting by tuning of hyperparameters. The SSD algorithm used pre-trained weights for the good generalization across the hieroglyphs. The pre-trained were achieved from the large datasets. The algorithms YOLOv5 and YOLOv8 show the framework of object detection that helps to achieve better mean Average Precision(mAP). These both models made sure that complex and overlapping hieroglyphs also be detected efficiently with the usage of low computational power. The Fast R-CNN model that is efficient in object detection has helped in making sure that proper and accurate detection is being done for detection and translating of hieroglyphs. All models were tested efficiently with different varieties of hieroglyphic images to make sure that hieroglyphic translation is done in efficient way by maintaining proper accuracy.

## 3.5 Model Evaluation

The evaluation of the models that have been trained are conducted to look after the effectiveness and robustness that translates and detects the hieroglyphs. This part has been done so that the insights of the performance, strengths and limitations could be known of the deep learning models that have been used. The other purpose of evaluating the models is also to make sure that the accurate results are achieved by analyzing the performance of the models on the separate dataset even on data that is not being seen. The evaluation is focused on the four models that are SSD, YOLOV5, YOLOV8 and Fast R-CNN. Where each model shows their own strength at the real time which made it efficient. SSD shows efficient results in object detection of different sizes, The strength of YOLOv5 and YOLOv8 is to detect object with high-speed using advance feature extraction techniques and provide better results. The strength of Fast R-CNN model is that it can handle the complex patterns of the hieroglyph even as some of them had overlapping effect. The key metric was analyzed in this evaluation part where every model's performance was checked that it should be meeting the requirements. The analysis was done as of metrices such as precision, recall, mean average precision (mAP), F1 score. Where precision helped in measuring of the accuracy of the positive predictions. Recall helped in identifying of the relevant hieroglyphs the mean average precision (mAP) showed performance among multiple thresholds and the F1 score gave the evaluation metric that was balanced.

# 4 Design Specification

The design specification incorporates about architecture and important functions of the implemented models, SSD, YOLOv5, YOLOv8, and Fast R-CNN. All the techniques that are used by each model for the optimization of the performance are explained. Each model carried out different architectures to handle the complexity of hieroglyphic images. Specifically models mechanism, networks for the extraction of features, loss function are focused. The framework of the models ensures that well designed and suitable to provide efficient results. Below provides detailed specifications and strength of each model

## 4.1 SSD

The architecture of the SSD model is designed in a manner to analyze the object detection, and it is done in real time. It does smooth feature detection and gives the best result. For which it does use MobileNetV2 that is also known as backbone when extracting the feature. In this study the resizing of the images has also been done as here it is resized to 300x300 pixel. It even ensured that it is compatible with the network. It even helps to detect if the object even varies in size by multi scale feature maps. The model works through layers as the objects are of different sizes, so it works in layers having the focus on scale. As detailed and broader layer could be introduced where detailed one could be used for small objects and the broader one could be used for big objects. The cross-entropy loss and smooth L1 loss are also used. Some data augmentation techniques have been also applied for more robust results. This all works in smooth way because of the model being lightweight.

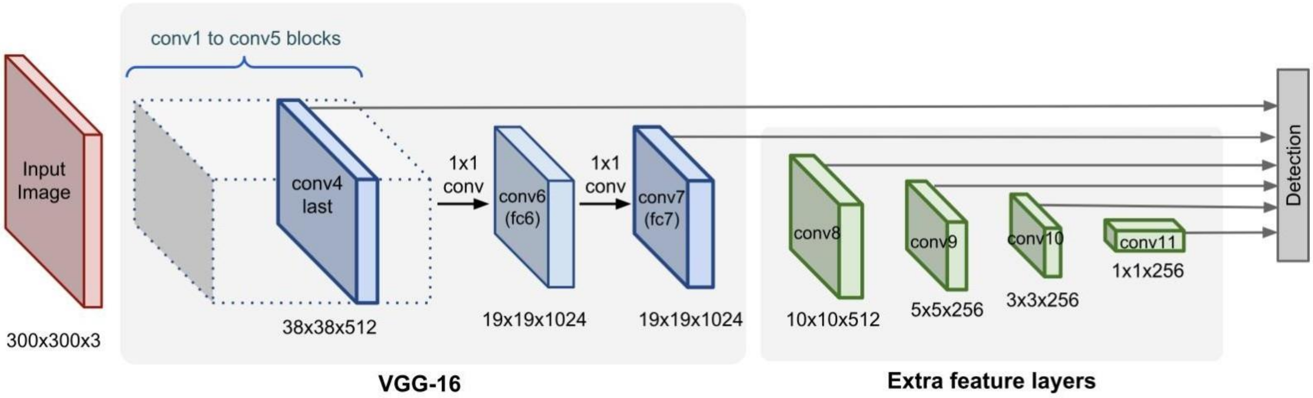



**Figure 5: SSD architecture with VGG-16 backbone and extra layers for multiscale feature detection.**

## 4.2 YOLOv5

YOLOv5 is also a deep learning model which fast and efficient. It is used to detect the objects from the image. It works in a way that it flows in one go which means that whole image is processed directly instead of doing it repeatedly. This is why it is also used in detecting the hieroglyphs. This model's architecture includes a cross-stage partial known as backbone, a path aggregation network as neck and a head. Where each one plays an important role. The backbone mainly does extraction of the features. It’s like this part identifies the patterns an image might have. The neck part whereas works for combining the information from every layer. The head is used to do the predictions of bounding box. It’s like finding out the score of objects, location of object and probability of the class, by using anchor boxes. Which helps to detect objects of different sizes efficiently. The model used image size for the detection of hieroglyphs is 320x320 pixels. For improved generalization data augmentation techniques were applied. This model’s architecture is essential for detecting object as it is a lightweight model.

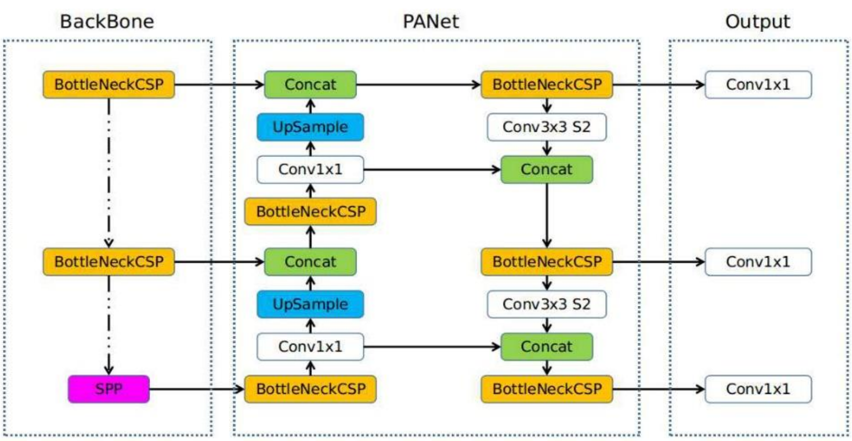



**Figure 6: YOLOv5 architecture with Backbone, PANet, and Output layers for object detection.**

## 4.3 YOLOv5

This model extracts the feature and provides framework that is optimized to detect the objects. This model is developed on the basis of predecessor which also updated the architecture for more better performance. The model is composed of backbone, head and neck framework helps to show better performance in detection of different complex objects within the image. the backbone of this architecture used CSPDarknet53 for feature extraction while maintaining the balance between computational cost and depth. the input images in this framework are resized to the dimensions of 640x640 pixels. The model includes multi-scale feature maps which are constructed through its neck region by using the network known as PANet which helps to enhance the feature extraction. This framework allows model to integrate spatial and contextual information from the different layers which makes it suitable for object detection over multiple scale. While training the anchor boxes are used. It’s the same as we do in YOLOv5 where we find out the score of objects, location of object and probability of the class. Prediction is done when the detection of hieroglyph is being done. The reason to use this model is for good accuracy and speed. This model uses techniques of data augmentation like random scaling, flipping and cropping which helps to improve the training dataset. The optimization of the model is done through ADAM optimizer having a learning rate of 0.001 which ensure to increase the accuracy. In deduction the technique known as non-maximum suppression is used for filtering out the detections that are unnecessary only keeping the important predictions. This model is suitable because of it giving best result in both ways as in computational and performance tasks too. As they help in detecting the hieroglyphic images of different sizes.

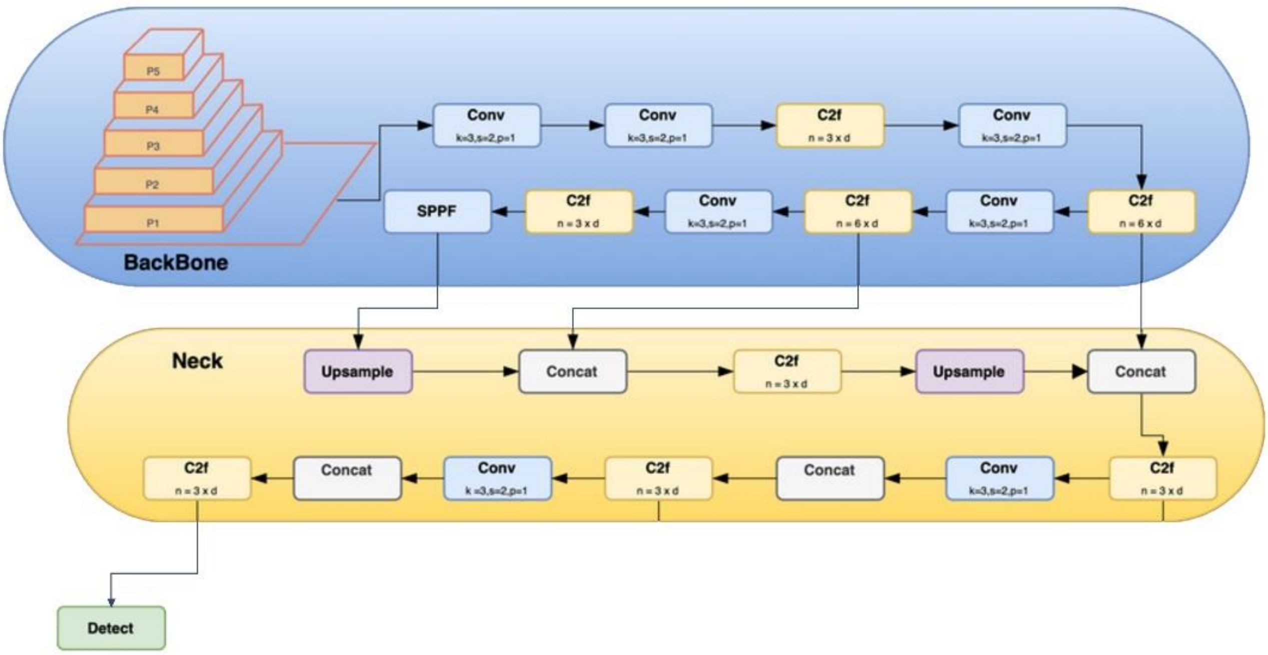



**Figure 7: Architecture of YOLOv8 showing the backbone, neck and head**

## 4.4 Fast R-CNN

This is an adopted library by Facebook Ai research that is used for detecting objects. Influencing the power of deep learning techniques to provide the effective performance in object detection tasks. The framework of this model is developed on the backbone of ResNet-50 with the integration of feature pyramid networks which is useful for feature extraction. The images are set to a standard resolution of 800x800 pixels for the maintenance and optimization. The Resnet-50 architecture the backbone of Fast R-CNN model extracts features from the hieroglyphic images and then proceed to the feature pyramid network for mapping multi- scale feature. these map helps to detect objects of different sizes. By using regional proposal network the model generate candidate object regions by investigating the areas that contain objects. These proposals are processed by using ROI align which helps to make sure that alignment of feature maps has done precisely with in the proposed regions which increases the accuracy of localization. the model uses two-stages for the detection. In first stage the RPN is utilized for the prediction of object region. In second stage the predicted region is classified into the categories. The training process combines the loss function to the optimize performance. the loss of classification reduces the error of categorical cross entropy for annotating correct labels. The regression loss of bounding box makes sure valid localization. The model is trained on the on the epochs of 100 and having a batch size of 16 to avoid overfitting problem and to enhance generalization data augmentation techniques are used like random, flipping, cropping. The Fast R-CNN is very optimized algorithm. Its efficiency in handling multi scale features and different complex training pipeline makes it suitable for detection of different complex object. using different techniques like FPN, RoI Align provides effective results for the detection of the Egyptian hieroglyphs.

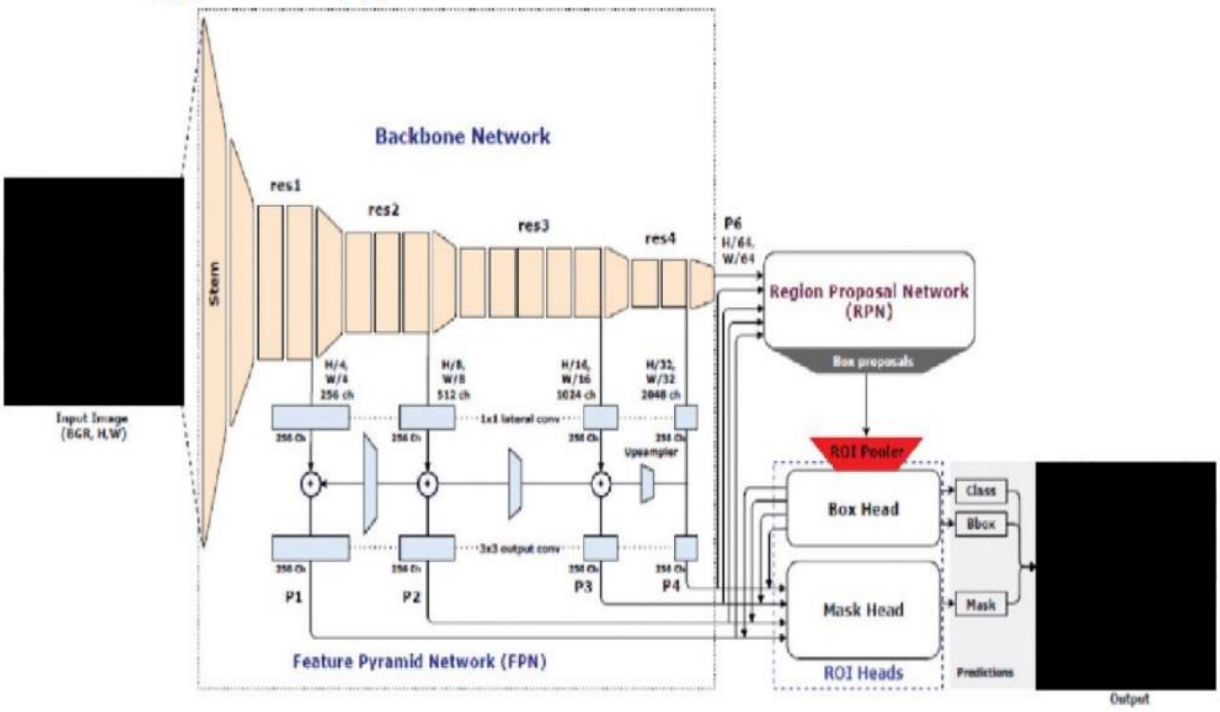



**Figure 8: The standalone model Fast R-CNN with Backbone Network, Feature pyramid network (FPN), Region proposal network (RPN), and ROI heads for object.**

# 5 Implementation

This study incorporates important steps of implementation for the object detection models such as SSD, YOLOv5, YOLOv8, Fast R-CNN. The process begins by dataset preparation, the annotated images in the dataset were divided into the sets of training, and validation. Dataset preprocessing was carried out to assure its closeness with the complex architecture of models. The data preprocessing involves steps like image reading, conversion of annotated images according to the specification of model's framework. Techniques of data augmentation were applied to improve the reliability of data. SSD model used TensorFlow for the training and deployment whereas remaining three models YOLOv5, YOLOv8, and FAST R-CNN used PyTorch. Each model was implemented according to its requirements. Pre-trained weights were used for the transfer learning. To load data custom training scripts were built, during training process hyperparameters and monitor performance were managed. Precision, loss and recall metrices were recorded to avoid overfitting problem and to detect accurate objects. GPU resources were used for the optimization of training process. Each model creates a bounding box along with the class labels for object detection. The detected objects were visualized through OpenCV library of python. To analyze the results of training process and for the visualization of key metrices Tensor Board and Matplotlib were used. Using validation data models were tested to make sure Egyptian hieroglyphs were detected accurately. All the libraries and tools that were used in the implementation allows flexible and adaptive way to integrate all four models. Helps to detect accurate objects within the image. This process shows that the model's efficiency in detecting the complex objects.

# 6 Evaluation

The all four models were implemented and evaluated on the basis of their performance for detection of Egyptian hieroglyphs. To check the performances of models the evaluation metrics like mAP, recall, F1 score and precision are measured. Different plots from training, validation accuracy and curve loss have been visualized to know the key findings from the results. The limitations and the strengths have also been considered so that properly objects could be detected no matter what sizes or shapes they hold.

## 6.1 Experiment 1 YOLOv5:

This model follows the framework that is efficient for identification in hieroglyphs by which it is implemented. The training was done using epochs as 100 of them were used. The batch size was of 64. The considered pixels were 320 x 320. Further efficiency was brought by the use of gradient flow which is effective in both low computational power and uses the CSP architecture. This helped in extracting features by giving the highest mAP that is 0.970. To improve feature fusion PANet is employed which enables model to detect object of different sizes and shapes effectively. The model shows the precision rate of 0.930 and recall rate of 0.970, indicates that model is capable of detection hieroglyphic images of different varieties, while making sure false positives are reduced. Anchor based detection techniques also influence the overall performance of model, by adjusting their sizes according to dimensions of the objects. The F1 score of 0.960 shows a significant balance between precision and recall. Below table shows the evaluation summary of the model

**Table 1: A table**

| Model Name | mAP | Precision | Recall | F1-Score |
|---|---|---|---|---|
| YOLO V5 | 0.970 | 0.930 | 0.970 | 0.960 |

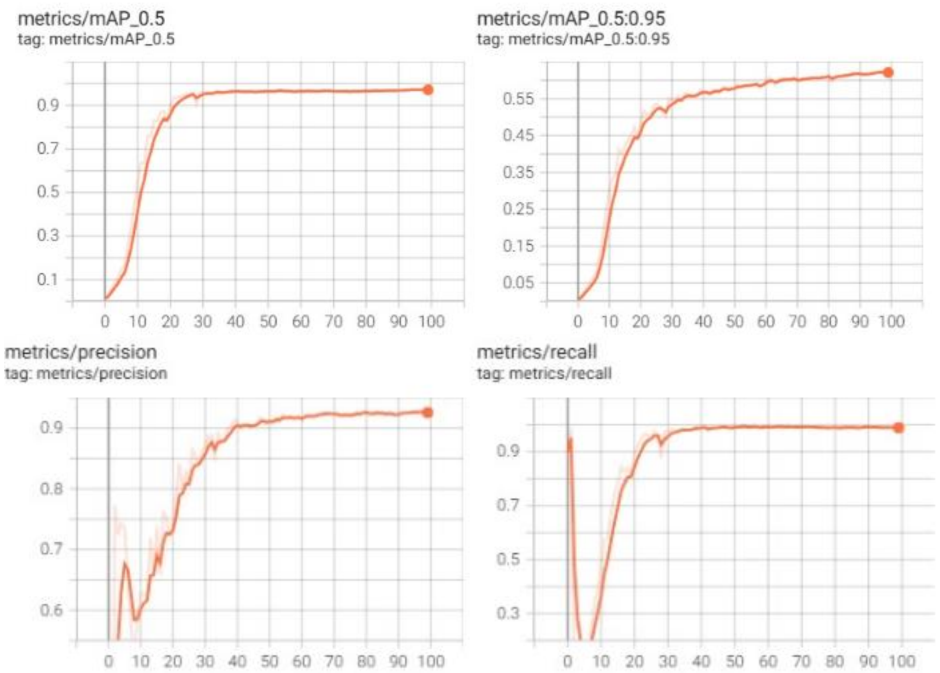



**Figure 9: Performance metrics of YOLOv5 during training, including mAP@0.5, mAP@0.5:0.results in the form of accuracy, precision, recall over epochs**

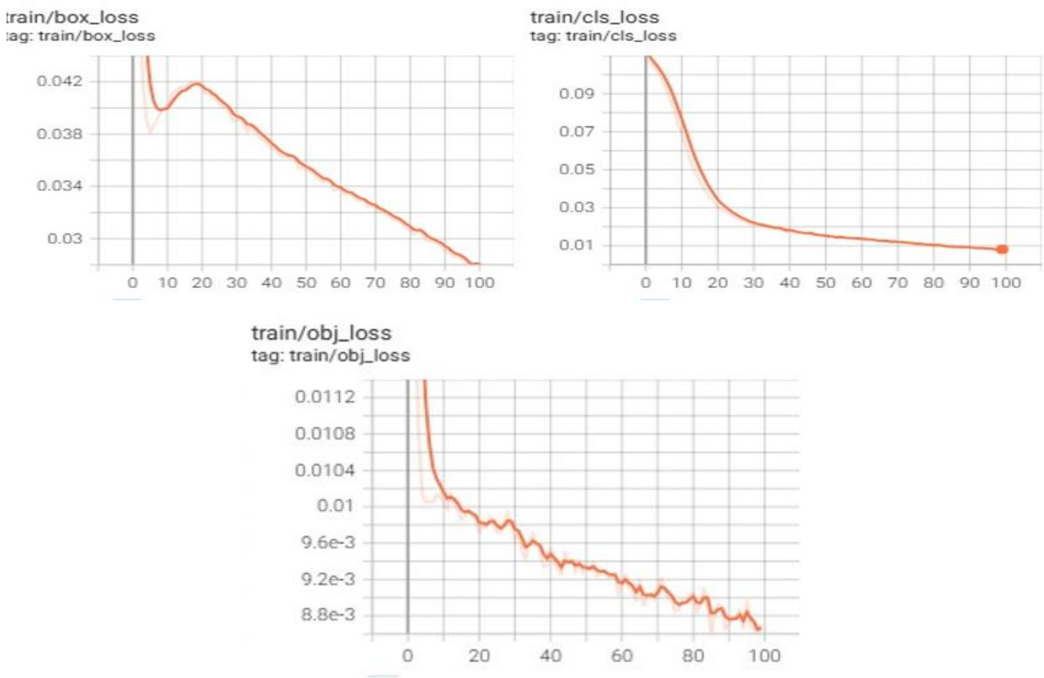



**Figure 10: Training loss trends for YOLOv5 showing box loss, classification loss, and abjectness loss over epochs**

The box loss and classification loss and abjectness loss show decrease suggesting accurate predictions of bounding box. Precision and recall showed continuous increase followed by stabilization of the 0.940 and 0.970. The mAP was 0.5 that represented sharp increase till 0.970 with the threshold of 0.5 showed higher accuracy. Whereas the graph having mAP 0.5:0.95. The mAP 0.5:0.95 at bottom increases to 0.55 did showed effectiveness of the model at the various thresholds.

## 6.2 Experiment 2 YOLOv8:

The model was implemented for its advanced framework and high efficiency in object detection task. For detection of Egyptian hieroglyph, it was trained on 100 epochs having the batch size of 54 and image size of 640x640. YOLOv8 is build upon the foundations of its previous architecture with the new advancement and modification in its backbone, enhance the detection ability of model. The model incorporates the anchor free detection approach helps to detect object localization more efficiently without depending on the predefined anchor boxes. The model achieved highest mAP of 0.975 as compare to other models, whereas the precision score of 0.943 shows efficiency in reduction of false positives. The model effectively detects the accurate object with high recall rate of 0.977. Balance between precision and recall is maintained by the F1 shows rate of 0.960. advanced architecture of model and its optimized learning techniques helps model to achieve high accuracy without prioritizing speed.

**Table 2: B table**

| Model Name | mAP | Precision | Recall | F1-Score |
|---|---|---|---|---|
| YOLOv8 | 0.975 | 0.943 | 0.977 | 0.960 |

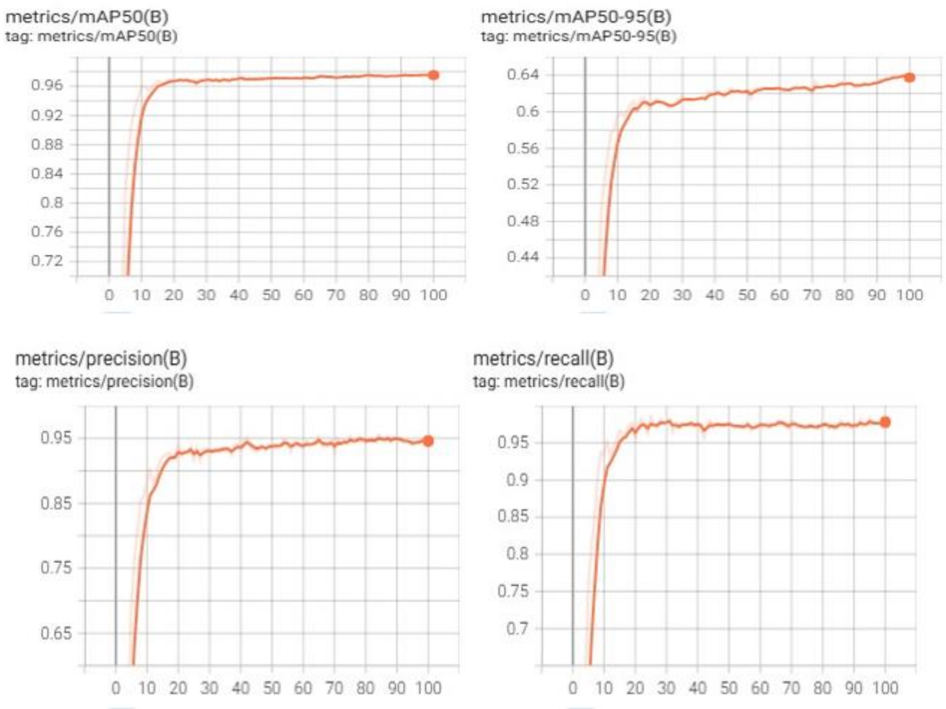



**Figure 11: Performance metrics of model during training, including mAP@50, mAP@50-95, precision, and recall trends across epochs**

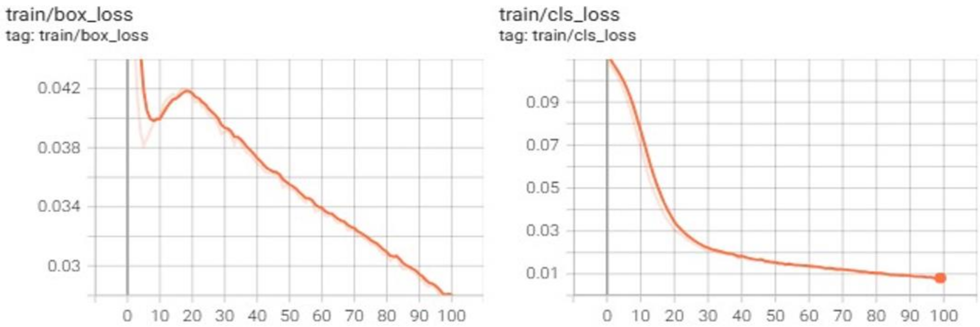



**Figure 12: Training loss curves for YOLOv8 showing box loss and classification loss reduction over epochs**

The mAP 0.5:0.95 metric shows rapid increase and stabilizes around 0.640 suggesting strong detection around different thresholds. Rapid increase of precision and recall and plateau at 0.950 shows that the model is able to reduce false positives and false negatives. The classification loss showing decrease is indicating prediction of class. The decline in box loss after some fluctuations is showing the proper regression of bounding box.

## 6.3 Experiment 3 Fast R-CNN:

The model uses region-based approach in its framework for the detection of hieroglyphs. The model was trained on 5000 epochs having the batch size of 8. The model employs a process of two stage detection, (RPN) helps to generate candidate regions and these regions are classified and their bounding box are refined in second stage. Whereas Feature pyramid network is combined with the architecture foe the extraction of multi scale feature and for accurate feature alignment RoI Align is used, mAP of 0.937 indicates model accurately classify and localize objects. The lower recall rate of 0.650 highlights limitation s in identification of accurate objects. The absence of precision and F1score shows that model face challenge to maintain the balance between false positives and false negatives. Besides absence of these values model shows strong performance in detection of hieroglyphs.

**Table 3: C table**

| Model Name | mAP | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fast R-CNN | 0.937 | N /A | 0.650 | N /A |

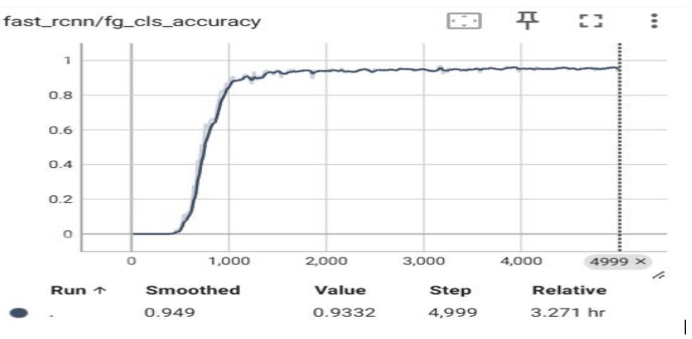



**Figure 13: Classification accuracy curve of Fast R-CNN showing foreground class accuracy over training steps**

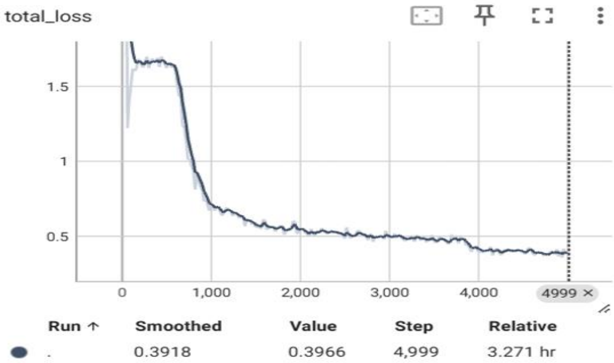



**Figure 14: Total loss curve of Fast R-CNN showing loss reduction over training steps**

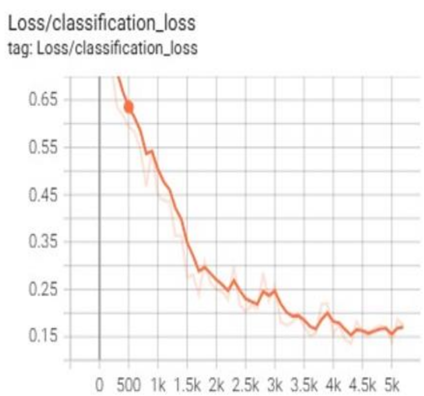
The top plot shows the total loss of steady decrease suggesting effective optimization which starts from high value and stabilized around 0.400. The bottom plot shows foreground object detection accuracy shows plateaus around 0.930 indicating model ability to detect accurate hieroglyphs.

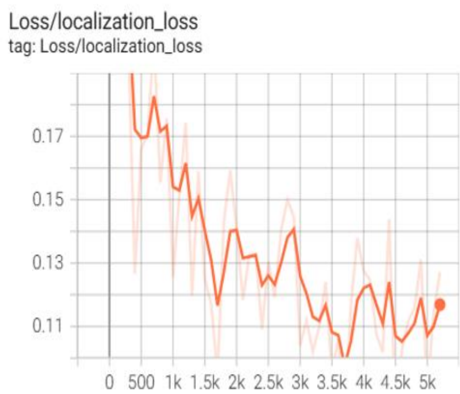
## 6.4 Experiment 4 SSD

The SSD was implemented and evaluated on the basis of its performance for detecting hieroglyphic image. This model uses MobileNetv2 which is light weight and maintains the balance between speed detection and computational power working as a backbone of this model. 5000 epochs were used for the training of the model while having the batch size of 16, ensures accurate object detection. The performance was evaluated by the mAP rate of 67.54, models show lowest mAP which is due to the limitations in its architecture. Single-shot detector uses single pass technique for the classification and prediction of objects location, while dealing with smaller and dense object the model compromises with precision, recall and F1score. Lowest number of feature maps were used in the multi-scale detection can also contribute to low mAP rate, impacting on the model performance to detect minor details efficiently, Below the table shows the evaluation parameters of the model.

**Table 4: D table**

| Model Name | mAP | Precision | Recall | F1-Score |
|---|---|---|---|---|
| SSD | 67.54 | N /A | N /A | N /A |



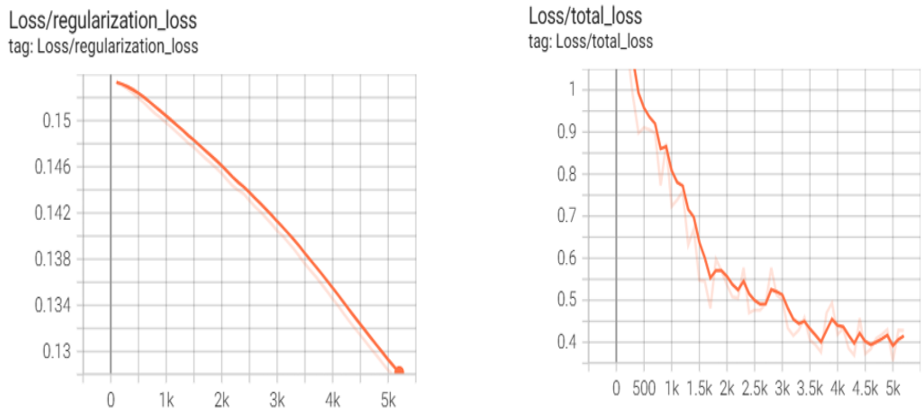



**Figure 15: Training loss curves for SSD model showing classification loss, localization loss, regularization loss, and total loss over training iterations.**

In the classification loss, the decrement could be seen which refers to improvement in accuracy prediction of the class. The total loss, also showing decrement suggests significant optimization and model convergence. The localization loss is indicating fluctuations but at initial stage as it is balanced afterwards. The regularization loss, also showing decline represents that no overfitting is seen in the model.

## 6.5 Model Comparison

The comparison of each model shows that SSD model achieves lowest value of mAP 67.54. Its single shot framework limits its detecting ability of small objects and gives priority to speed over precision. The Fast R-CNN uses region-based approach for the detection and shows significant mean average precision score of 0.937, YOLOv5 also shows excellent results, utilized its advance backbone architecture and modified multi-scale feature extraction and achieved 0.970. Although YOLOv8 achieved higher mAP of 0.975 among all the models by employing anchor free techniques which shows its detection capability with high precision of 0.943 and recall value of 0.977.

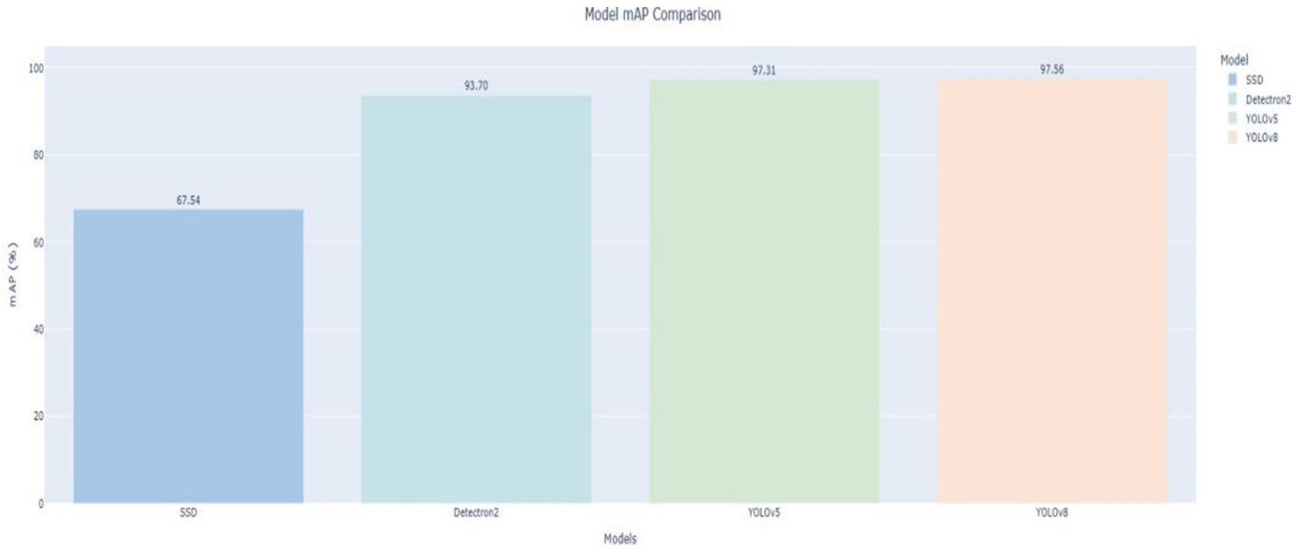



**Figure 16: Comparison of mAP performance across SSD, Fast R-CNN, YOLOv5, and YOLOv8 models.**

## 6.6 Discussion

This study discussion includes the strengths, functioning and limitations of the four models that are SSD, YOLOv5, YOLOv8 and Fast R-CNN. It also includes the key metrics that are recall, precision, F1 score and mean average precision. In complex scenarios, the limitation were highlighted as the SSD model was limited due to its inability of detecting the hieroglyphs that are small. The YOLOv5 and YOLOv8 whereas achieved good and high mean average precision when it is compared to SSD. The model that has out performed was YOLOv8 as it got the highest mean average precision. As YOLOv8 is more efficient when working with complex scenarios if it is checked with YOLOv5 but this comes with lowering the speed. The Fast R-CNN was seen as the most reliable model including the multiple scale features. It does have the ability to function in the backgrounds which are complex in nature. Even after using techniques like data augmentation the problem stays still and these challenges can also so be seen because of the predetermined size that is anchored to the boxes in the models like SSD. The generational capability can also be put emphasis on when it varies in a textured and a plain background. All these challenges can easily be addressed through preprocessing and by using various strategies in future. It is seen that YOLOV8 has outperformed as it has got the best result among all the models that have been evaluated. As the mAP achieved by YOLOV8 is around 0.975 having the precision value of 0.943, recall value of 0.977 and F1 score value being 0.960 which is highest then all other models.

# 7 Conclusion and Future Work

This study uses complex deep learning models and provides a powerful framework for the detection of Egyptian hieroglyph. The models SSD, YOLOv5, YOLOv8 and FAST R-CNN were implemented and their performance for the detection of Egyptian hieroglyph were evaluated through an evaluation metrics. This research provides an accurate understanding of the research objectives and present a critical review of the related articles in the section of literature review. The methodology parts discuss about all the various steps that were fulfilled during the execution. It briefly discusses about the dataset preparation, uncover the steps of EDA, explains about model training and its overall performance. The word-based architecture of the models is explained in the design specification part. Each model's performance was evaluated and explained along with their key findings and limitations, the results of the evaluation shows that YOLOv8 is considered as the best model having mAP of 0.975 and high precision and recall of 0.943 and 0.977 and F1 score of 0.960. The SSD model comprises with precision, recall, and F1 score and shows mAP of 67.54, the FAST R-CNN shows absence of precision and recall score and achieved mAP of 0.937 and recall of 0.650, the YOLOv5 shows the mAP of 0.973, precision of 0.926 and recall of 0.987 and F1 score of 0.960. Techniques like data augmentation and process training have also been used so that more efficient performance could be observed in the models.

Future work can emphasize to Convert the best model into the TF-Lite and deploy it on mobile devices and integrated into the flutter and android applications along with web app. which enables to provide a user-friendly environment for tourist, which allows detection and translation of hieroglyph in real time. The development of these application provides a way to preserve cultural heritage and helps many tourists to explore the culture without hiring any expensive guides.

# References

Barucci, A., Cucci, C., Franci, M., Loschiavo, M., & Argenti, F. (2021) 'A deep learning approach to ancient Egyptian hieroglyphs classification', IEEE Access, 9, pp. 123438–123447.

Corsini, M., Ferdani, D., Kuijper, A., & Kutlu, H. (2024) 'Unsupervised detection and localization of Egyptian hieroglyphs'.

Elnabawy, R., Elias, R., Salem, M. A. M., & Abdennadher, S. (2021) 'Extending Gardiner’s code for hieroglyphic recognition and English mapping', Multimedia Tools and Applications, 80, pp. 3391–3408.

Franken, M., & van Gemert, J. C. (2013) 'Automatic Egyptian hieroglyph recognition by retrieving images as texts', in Proceedings of the 21st ACM International Conference on Multimedia, pp. 765–768.

Cucci, C., Guidi, T., Picollo, M., Stefani, L., Python, L., Argenti, F., & Barucci, A. (2024) 'Hyperspectral imaging and convolutional neural networks for augmented documentation of ancient Egyptian artefacts', Heritage Science, 12(1), p. 75.

Moustafa, R., Hesham, F., Hussein, S., Amr, B., Refaat, S., Shorim, N., & Ghanim, T. M. (2022) 'Hieroglyphs language translator using deep learning techniques (Scriba)', in 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), pp. 125–132. IEEE.

Verma, S., Gupta, N., Anil, B. C., & Chauhan, R. (2022) 'A novel framework for ancient text translation using artificial intelligence', ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 11(4), pp. 411–425.

Malhotra, R., & Addis, M. T. (2023) 'End-to-end historical handwritten Ethiopic text recognition using deep learning', IEEE Access.

Jindal, A., & Ghosh, R. (2024) 'A hybrid deep learning model to recognize handwritten characters in ancient documents in Devanagari and Maithili scripts', Multimedia Tools and Applications, 83(3), pp. 8389–8412.

Narang, S. R., Kumar, M., & Jindal, M. K. (2021) 'DeepNetDevanagari: A deep learning model for Devanagari ancient character recognition', Multimedia Tools and Applications, 80, pp. 20671–20686.

Alshahrani, B. N., & Alghamdi, W. Y. (2024) 'A deep learning approach to convert handwritten Arabic text to digital form', International Journal of Advanced Computer Science & Applications, 15(5).

Bennour, A., Boudraa, M., Siddiqi, I., Al-Sarem, M., Al-Shabi, M., & Ghabban, F. (2024) 'A deep learning framework for historical manuscripts writer identification using data-driven features', Multimedia Tools and Applications, pp. 1–27.

Umamaheswari, S., Gowtham, G., & Harikumar, K. (2024) 'Bridging the linguistic gap: A deep learning-based image-to-text converter for ancient Tamil with web interface', Automatic Speech Recognition and Translation for Low Resource Languages, pp. 213–241.

Kadam, D., Jadhav, A., Govilkar, P., Bhosale, Y., Jadhav, S., & Gamre, S. (2024) 'Unveiling the past: A holistic approach to rescuing historical texts through advanced image analysis and AI'.

Preethi, P., & Mamatha, H. R. (2023) 'Region-based convolutional neural network for segmenting text in epigraphical images', in Artificial Intelligence and Applications, 1(2), pp. 119–127.

Sarala, A., & Vinothkumar, C. (2024) 'Unlocking the past: Decoding historic inscriptions with the power of convolutional recurrent neural networks', in 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), pp. 1–7. IEEE.

Haliassos, A., Barmpoutis, P., Stathaki, T., Quirke, S., & Constantinides, A. (2020) 'Classification and detection of symbols in ancient papyri', Visual Computing for Cultural Heritage, pp. 121–140.

Hidayat, A. A., Purwandari, K., Cenggoro, T. W., & Pardamean, B. (2021) 'A convolutional neural network-based ancient Sundanese character classifier with data augmentation', Procedia Computer Science, 179, pp. 195–201.

Islam, M. A., & Iacob, I. E. (2023) 'Manuscripts character recognition using machine learning and deep learning', Modelling, 4(2), pp. 168–188.

Gunasekara, S., Lafir, M. H., Dulaj, C., Haputhanthri, L., & Alwis, D. (2024) 'Deep learning-powered mobile app for early Brahmi script decipherment in Sri Lanka', in 2024 International Research Conference on Smart Computing and Systems Engineering (SCSE), 7, pp. 1–6. IEEE.