# Khmer Text Detection using YOLO
# A Computer Vision Approach

**Ms. Siv Sreynoch‡ Mr. Tet Davann‡ Mr. Ngang Puthsabbos‡ Mr. Sok Yongyi‡**

School of Digital Engineering
Cambodia Academy of Digital Technology, Phnom Penh, Cambodia
{sreynoch.siv, davann.tet, puthsabbos.ngang, yongyi.sok}@student.cadt.edu.kh

## Abstract

This paper presents a deep-learning approach for detecting Khmer text using the YOLO object detection model. The project aims to overcome challenges in detecting Khmer script, which has unique characteristics compared to other languages. We describe the collection, annotation, and preparation of a custom Khmer text dataset, followed by the training of a YOLO model on this dataset. Experimental results show that the model achieved an mAP of 87%, with promising performance in detecting Khmer text. This research provides a solid foundation for future developments in Khmer language processing.

**Keywords:** Khmer Text Detection, Computer Vision, YOLOv5, YOLOv8, YOLOv9, YOLOv11

## 1 Introduction

Khmer text detection is an important task in computer vision, as it has numerous applications in areas such as document analysis, image retrieval, and scene understanding. However, detecting Khmer text in images is a challenging task due to the diversity of fonts, sizes, and orientations of Khmer text in natural images, as well as varying lighting and environmental conditions.

## 2 Dataset

We collected a dataset of images containing Khmer text by manually capturing images of banners, signs, or objects containing Khmer text. The dataset consists of 6599 images, with 5324 images used for training, 496 images used for testing, and 779 images used for validation.

We used the VGG tool to annotate images. To prepare the dataset for training, we utilized Roboflow to convert the bounding boxes of the images into polygon format suitable for YOLO. Additionally, we performed image preprocessing, which included resizing the images to 640x640 pixels and transforming them to grayscale.

## 3 Experiment setup

We trained the YOLOv5(yolov5x6u), YOLOv8(yolov8n), YOLOv9(yolov9c), and YOLOv11(yolo11n) models using the collected dataset. We used the Adam optimizer with a learning rate of 0.002 and a batch size of 16. We trained the models for 100 epochs and evaluated their performance using precision recall, mAP50, mAP50-95. We just use the model's default configuration with an epoch set to 100.

## 4 Evaluation and Result

In this section, we evaluate the performance of different versions of the YOLO model (YOLO v3, YOLO v4, YOLO v5, and YOLO v8) on the Khmer text detection task. The evaluation is conducted using several standard performance metrics, including mean Average Precision (mAP), precision, recall. The evaluation results for each YOLO model version are summarized in **Table 1**.

Table 1: Evaluation Results for YOLO Model

| Model | **Precision** | **Recall** | **mAP 50** | **mAP 50-95** | **Fitness** |
|---|---|---|---|---|---|
| **YOLOv5** | 0.91 | 0.61 | 0.68 | 0.46 | 0.48 |
| **YOLOv8** | 0.89 | 0.62 | 0.68 | 0.46 | 0.48 |
| **YOLOv9c** | 0.89 | 0.63 | 0.68 | 0.47 | 0.49 |
| **YOLOv11** | 0.86 | 0.62 | 0.68 | 0.47 | 0.49 |

Discussion of Results:

- **YOLOv5** showed the highest precision (0.91), indicating strong detection accuracy but lower recall (0.61), suggesting it may miss some instances of Khmer text.
- **YOLOv8** balanced precision (0.89) and recall (0.62), making it slightly more reliable than YOLOv5 in detecting all instances.
- **YOLOv9** achieved the highest recall (0.63) and better mAP@50-95 (0.47), showing improved detection capability for diverse cases.
- **YOLOv11** had slightly lower precision (0.86) compared to other models but performed comparably in recall and mAP metrics, with a fitness score of 0.49.These results indicate that while YOLOv5 excels in precise detection, YOLOv9 and YOLOv11 provide a better balance between detection accuracy and recall, making them more suitable for comprehensive Khmer text detection tasks.

## 5 Discussion and Conclusion

The results of our experiments show that the YOLOv9 model is the best-performing model for Khmer text detection. The high precision and recall rates achieved by the YOLOv9 model demonstrate its effectiveness in detecting Khmer text in images. However, there are some limitations to our study. The dataset used in this study is relatively small, and the models may not generalize well compared to other datasets. Additionally, the models may not perform well in cases where the Khmer text is distorted or occluded. Future work should focus on collecting a larger dataset and improving the models to handle cases where the Khmer text is distorted or occluded. Our code repository is available at: https://github.com/Nochsiv/Khmer-Text-Dection-Computer-Vision-Group4.git

**Acknowledgment**

We would like to express our deepest gratitude to Ms. Nom Vannkhim and Mr. Him Soklong, our lecturers, for their invaluable guidance, encouragement, and support throughout this project. Their expertise and insights have been instrumental in shaping our understanding of Khmer text detection and deep learning methodologies.

We are also grateful to the School of Computer Science for providing us with the resources and opportunities to work on this project. Special thanks to our classmates and team members for their dedication, collaboration, and hard work in overcoming the challenges we encountered.

Finally, we extend our heartfelt thanks to our families and friends for their unwavering support and encouragement, which motivated us to complete this project successfully.