



សាកលវិទ្យាល័យភូមិន្ទភ្នំពេញ

ROYAL UNIVERSITY OF PHNOM PENH

**ការស្រាវជ្រាវវិធីសាស្ត្រថ្មី សម្រាប់កំណត់សម្គាល់អត្ថបទអក្សរខ្មែរ
ដែលសរសេរដោយដៃ នឹងបានប្រើប្រាស់ស្ថាបត្យកម្ម Craft ជាមួយនឹង TrOCR**

**A novel End-to-End approach to Khmer Handwritten Text
Recognition using Craft with TrOCR Architecture**

Mr. Vitou Soy

A Thesis

In Partial Fulfilment of the Requirement for the Degree of
Bachelor of Engineering in Information-Technology-Engineering
Academic Year 2022 - 2025

Examination committee: Mr. Sokchea Kor (Advisor)
Mr. Chanpiseth Chap (committee)
Mrs. Daly Chea (committee)
Dr.

មូលន័យសង្ខេប

ដោយសារតែកង្វះឯកសារ ការចងក្រងអក្សរខ្មែរ នៅមានកម្រិតហេតុដូចនេះហើយ ទើបមានការខ្វះខាត ជាខ្លាំងទៅលើ ការស្រាវជ្រាវអក្សរសាស្ត្រខ្មែរ ជាពិសេស គឺការសិក្សាទៅលើ ការកំណត់សម្គាល់ អត្ថបទអក្សរសាស្ត្រខ្មែរ ដែលបានសរសេរដោយដៃ មូលហេតុចម្បងដែលវាពិបាកនោះគឺ ការបង្កើត ឯកសារសិប្បនិម្មិត (synthetic data) គឺវាមានភាពស្មុគស្មាញ ដើម្បីទុកសម្រាប់ស្រាវជ្រាវទៅលើ ប្រព័ន្ធ ទាំងអស់នោះ។ នៅក្នុង អត្ថបទស្រាវជ្រាវមួយនេះ ខ្ញុំនឹងប្រើប្រាស់វិធីសាស្ត្រដែល មានសក្តានុពលសម្រាប់ ការកំណត់សម្គាល់ អក្សរសាស្ត្រខ្មែរដែលបានសរសេរចេញពីដៃផ្ទាល់

Abstract

A novel approach to Khmer handwritten text recognition is proposed using a fine-tuned TrOCR transformer model. The system leverages large-scale multilingual pretraining and supervised fine-tuning to handle complex character structures in Khmer script. This thesis presents the dataset preparation, model architecture, training methodology, and performance evaluation using standard benchmarks. The results show significant improvements in recognition accuracy compared to traditional OCR techniques.

Contents

| | | |
|---|--------------|---|
| 1 | Introduction | 3 |
|---|--------------|---|

Chapter 1

Introduction

This is the introduction chapter.