MINISTRY OF EDUCATION AND TRAINING

HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY AND EDUCATION

FACULTY OF HIGH QUALITY TRAINING

FINAL-TERM REPORT (SUMMARY)

ARTIFICIAL INTELLIGENCE

LANGUAGE DETECTION (CNN MODEL)

Lecturer: MSc Nguyen Truong Thinh

SVE: Nguyen Ngoc Quy

Major: Mechatronics Engineering

ID: 20146056

Ho Chi Minh City, May 26, 2023

LANGUAGE DETECTION

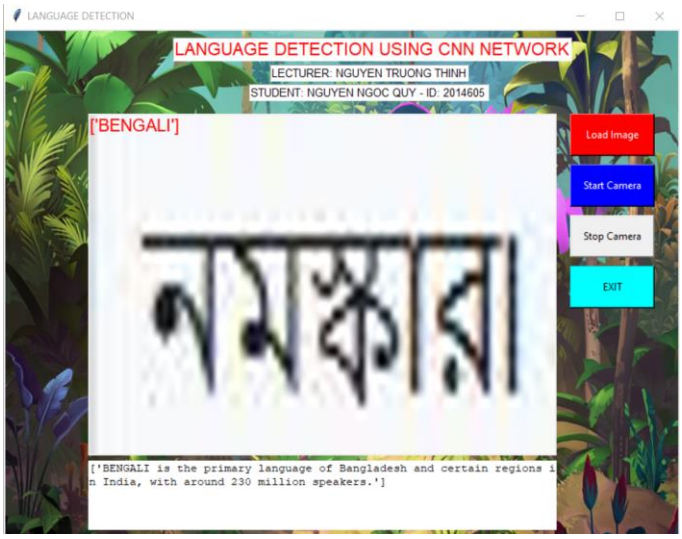
This report introduces the important role of language detection in text classification and cross-cultural communication. Using Convolutional Neural Networks (CNNs), we propose building an intelligent language detection system capable of accurately identifying and classifying different languages based on textual inputs. By combining theoretical knowledge with practical implementation, our goal is to develop a user-friendly application that allows users to easily identify the language of any text, regardless of its origin or complexity.

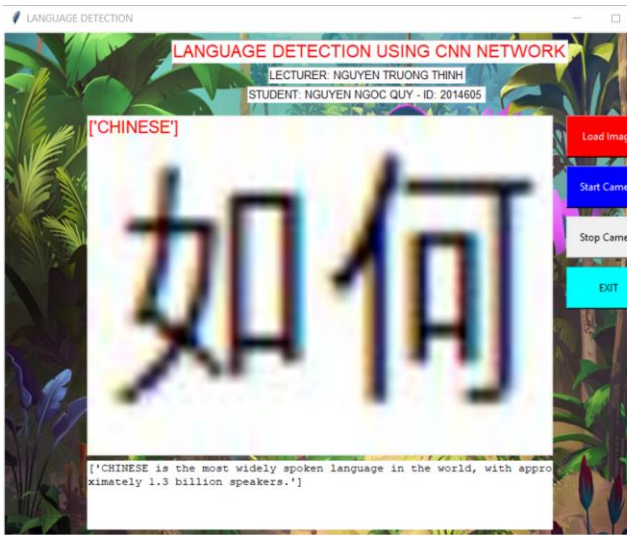
Our research methodology includes data loading, data splitting into training and test sets, CNN model construction, data augmentation, model training and evaluation. By applying these steps, our language detection system effectively classifies text into different languages with accuracy, enabling various language-related applications.

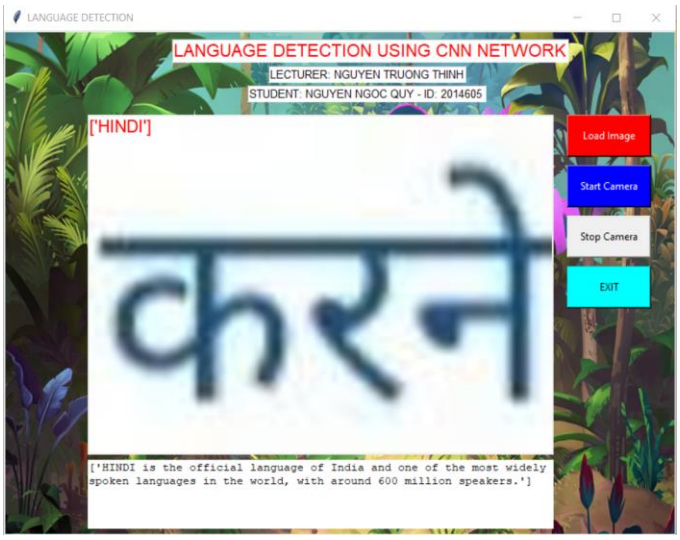
This report holds significant importance in the development of natural language processing technology. By building an accurate and user-friendly language detection system, we contribute to this field and open up potential future advancements.

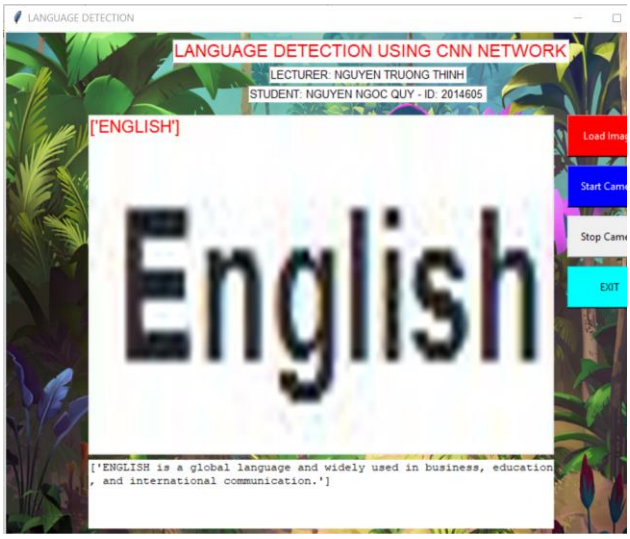
RESULTS OF LANGUAGE DETECTION

1. **Results of detection language with thinker load data.**




LANGUAGE DETECTION
LANGUAGE DETECTION USING CNN NETWORK
LECTURER: NGUYEN TRUONG THINH
STUDENT: NGUYEN NGOC QUY - ID: 2014605
['CHINESE']
Load Image
Start Camera
Stop Camera
EXIT
['CHINESE is the most widely spoken language in the world, with approximately 1.3 billion speakers.']


LANGUAGE DETECTION
LANGUAGE DETECTION USING CNN NETWORK
LECTURER: NGUYEN TRUONG THINH
STUDENT: NGUYEN NGOC QUY - ID: 2014605
['HINDI']
Load Image
Start Camera
Stop Camera
EXIT
['HINDI is the official language of India and one of the most widely spoken languages in the world, with around 600 million speakers.']


LANGUAGE DETECTION
LANGUAGE DETECTION USING CNN NETWORK
LECTURER: NGUYEN TRUONG THINH
STUDENT: NGUYEN NGOC QUY - ID: 2014605
['ENGLISH']
Load Image
Start Camera
Stop Camera
EXIT
['ENGLISH is a global language and widely used in business, education, and international communication.']


LANGUAGE DETECTION
LANGUAGE DETECTION USING CNN NETWORK
LECTURER: NGUYEN TRUONG THINH
STUDENT: NGUYEN NGOC QUY - ID: 2014605
['INDONESIAN']
Load Image
Start Camera
Stop Camera
EXIT
['INDONESIAN is the official language of Indonesia and widely used in communication, education, and media.']


LANGUAGE DETECTION
LANGUAGE DETECTION USING CNN NETWORK
LECTURER: NGUYEN TRUONG THINH
STUDENT: NGUYEN NGOC QUY - ID: 2014605
['FRENCH']
Load Image
Start Camera
Stop Camera
EXIT
['FRENCH is the official language of many countries and widely used in culture, arts, and legal documents.']


LANGUAGE DETECTION
LANGUAGE DETECTION USING CNN NETWORK
LECTURER: NGUYEN TRUONG THINH
STUDENT: NGUYEN NGOC QUY - ID: 2014605
['ITALIAN']
Load Image
Start Camera
Stop Camera
EXIT
['ITALIAN is the official language of Italy and considered one of the most popular Romance languages.']

2. Results of detection language with thinker RealTime.