# ANN final project report

## AI powered signboard translation

By: Mohamed Salama 211001298

Youssef Yacoub

Ahmed Hamdy

Seif hashem  
adham

## Introduction

This project focuses on developing a deep learning-based system that performs real-time translation of signboards from one language to another. By combining computer vision (CV) and natural language processing (NLP), the system detects text in signboards, translates it, and overlays the translated version back onto the image. This technology is particularly useful for travelers, tourists, and language learners navigating unfamiliar regions.

## Objective

To create a pipeline that:

* Detects text from images (signboards)
* Translates the extracted text into a target language
* Overlays the translated text on the original image in a visually coherent manner

### **3. Methodology**

#### **3.1 Pipeline Overview**

1. **Text Detection:**
   * Dataset: COCO-Text, TextOCR.
2. **Text Recognition (OCR):**
   * Applied CRNN (Convolutional Recurrent Neural Network) for sequence recognition.
3. **Translation:**
   * Integrated a transformer-based translation model (e.g., MarianMT or M2M100).
4. **Text Overlay:**
   * Detected regions are masked and redrawn using OpenCV.
   * Translated text is rendered in-place using appropriate fonts and styling.

### *3.2 contribution*

**1. OCR:**

We experimented with multiple OCR engines to identify the most suitable one for our signboard translation pipeline:

* **Tesseract OCR**: Lightweight and open-source, but struggled with complex fonts, stylized text, and images captured in low lighting.
* **EasyOCR**: Provided better multilingual support and handled some challenging fonts better than Tesseract. However, it was slower and had difficulty with rotated or curved text.
* **PaddleOCR**: Offered a good balance between speed, accuracy, and ease of integration. It supports multiple languages and works well in real-time settings.
* **Google Vision OCR**: Achieved the highest text recognition accuracy among all tools tested. However, it requires an active internet connection and is a paid service, which limits its feasibility for real-time, offline deployment.

Based on our evaluation, **PaddleOCR** was chosen as the primary OCR tool for this project due to its strong performance and suitability for on-device usage.

**2. Translation Model:**

* **Marefa Translation Model**: An Arabic-focused transformer model that delivered strong results for Arabic-English translation tasks. However, it had limited generalizability to other languages and lacked customization flexibility.
* **Helsinki-NLP (MarianMT)**: A multilingual translation model with broad language support and open-source accessibility. It performed well out of the box and allowed us to fine-tune on domain-specific data.

After evaluating both, we chose **Helsinki-NLP** as the core translation model due to its multilingual capabilities and extensibility. We further fine-tuned it using a subset of signboard-related parallel text data to improve contextual translation accuracy for short, fragmented, or informal sign text.

### *3.3 results:*

The Helsinki model was fine tuned for 3 epochs, reducing training loss from 0.0733 to 0.0001, indicating strong convergence and suggests improved translation quality on training data.

Evaluating OCR systems can be challenging due to the lack of standardized, ground-truth annotations for real-world signboard images. Unlike translation models, OCR models do not always lend themselves to traditional accuracy metrics, especially when working with multilingual, varied, and noisy visual data.

As such, instead of relying on numerical metrics, we qualitatively assessed the performance of different OCR models (Tesseract, EasyOCR, PaddleOCR, Google Vision) by observing their ability to correctly detect and extract text from real signboard images.Below is a representative example of OCR output using our chosen model (**PaddleOCR**)

A person in a room with a video game controller

AI-generated content may be incorrect.

This visual evidence demonstrates PaddleOCR’s high accuracy in multilingual, real-world scenarios — a critical requirement for our real-time signboard translation pipeline.

To write a solid **"Comparison with Related Work"** section for your **AI-Powered Signboard Translation** project, you should briefly highlight how your system performs or innovates compared to previous solutions across **OCR**, **Translation**, and **End-to-End Integration**.

Here’s a structured paragraph you can paste into Word and tweak as needed:

## 4.Comparison with Related Work

Several existing solutions address parts of the signboard translation problem, but few offer a complete, real-time pipeline. Traditional OCR systems like Tesseract and EasyOCR offer basic text extraction but struggle with complex or multilingual signboards. Cloud-based models such as Google Vision OCR deliver superior performance but rely on internet access and incur usage costs, limiting their offline applicability. In contrast, our choice of PaddleOCR balances accuracy and efficiency while remaining fully open-source and offline-capable.

Below is the output of tesseract OCR, completely missing most of the text on screen, also before adjusting the image enhancement filters.

A street sign on a pole

AI-generated content may be incorrect.

For translation, we compared Helsinki-NLP, which we fine-tuned, and Marefa, a strong Arabic-English model. While both yielded readable translations, our fine-tuned Helsinki model showed smoother domain adaptation and lower loss over time, suggesting better generalization.

Unlike isolated solutions, our system integrates OCR, translation, and overlay rendering into one cohesive pipeline, making it suitable for real-time use cases such as AR navigation or travel assistance. This end-to-end integration distinguishes our approach from most academic or commercial tools that only handle a single component.

Below is the output of the translation model previously

A screenshot of a computer

AI-generated content may be incorrect.

## Challenges faced

**OCR Noise:** Detected text was often incomplete or incorrect, requiring pre- and post-processing.

**Overlay Alignment:** Ensuring text fits naturally into the image region across various shapes and lighting conditions.

**Multilingual Support:** Training translation models for low-resource languages required data augmentation techniques.

**Translation Model Adaptation**: While Marefa was strong for Arabic-specific tasks, its scope was limited for our multilingual goals. Helsinki-NLP provided a broader foundation but required fine-tuning to handle short and non-standard signboard phrases effectively. Preparing appropriate training data for fine-tuning and preventing overfitting on small datasets were key challenges.

**OCR Engine Trade-offs**: One of the initial challenges was choosing the right OCR tool. While Google Vision OCR delivered the best recognition accuracy, its dependency on internet connectivity and paid API usage made it impractical for offline and scalable real-time use. Tesseract and EasyOCR were viable but had limitations in terms of accuracy and performance. PaddleOCR emerged as the best option, balancing recognition quality with offline functionality.

## Conclusion

We successfully developed a deep learning-based pipeline for real-time signboard translation that performs detection, recognition, translation, and text overlay efficiently. The improvements in OCR accuracy, translation quality, and runtime demonstrate the viability of this solution for on-device use in real-world applications.