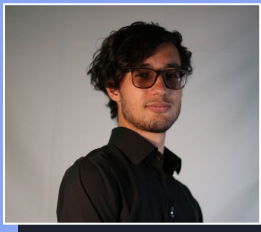# CS Capstone Design - AR Object Detection and Text Recognition for Language Learning
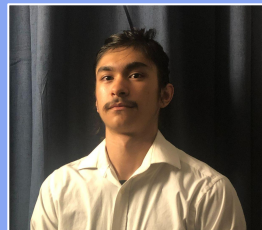
Team LangLens

Team LangLens

Stefan Mihailovic
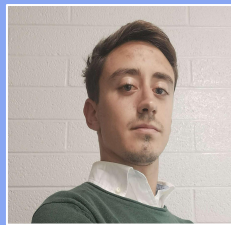Team Lead
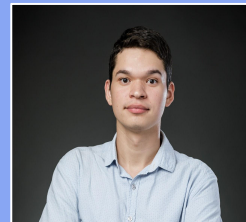
Brian Ruiz
Recorder

Sami Tanquary
Architect / Web Developer

Kyle Young
Customer Communicator

Daniel Navarrete
Release Manager

Italo Santos
Faculty Mentor

Our Client

Dr. Okim Kang

- ❖ **Professor in the Department of English at Northern Arizona University**
- ❖ **Specialized knowledge in Applied Linguistics and Teaching English as a Second Language (TESL)**
- ❖ **Received many awards and honors for her contributions to language pedagogy research**
- ❖ **Worked on developing systems / apps dedicated to linguistics in the past**

NORTHERN ARIZONA UNIVERSITY

Problem Statement

Lack of **free**, **accessible** tools for language learners that utilize **both** object detection and text recognition

Current learning applications **do not focus** on the key elements of word learning: *meaning, usage,* and *form*

Current tools are **challenging** for those who struggle with text or are inexperienced in navigating a foreign language

Our Solution

Free

Free-to-use and accessible to anyone

Mobile

Web-app optimized for mobile devices

Modern

Employs both object AND text recognition capabilities

Word Learning

Covers key aspects of word learning: *meaning, usage,* and *form*

LangLens Mobile Version

Native Language Toggle

Select Language ▼

LangLens

A web-based mobile app for language learning implementing AR technology.

Choose a Detection Mode to Start

Object Mode 🚗   Text Mode 📖

What is LangLens?

LangLens is a small team of 5 members where we are set to create a language learning tool that will be free for everyone around the globe to use. We understand that language is an important skill that everyone has the potential to learn and is something everyone should

Homepage GUI

Requirement Acquisition Plan

Weekly Team Meetings

1. Review Goals
Discussion of MVP, review development timeline, and discuss next MVP requirement

2. Rate Feasibility
what is the problem and *how* will we solve it with this feature? Is it technically possible?

3. Refine
Check if user story is small enough for a single sprint or needs to be split into multiple

4. Acceptance
Discuss the conditions of satisfaction. What is the desired outcome / how can we test it?

5. Time Estimation
Discuss how long this requirement is estimated to take to develop

6. Assign New Feature
Only if previous feature is fully refined and satisfactory / ready

Requirements Review

Key Requirements

- ❏ Toggle between object detection and text recognition modes
- ❏ Target language selector (MVP: Spanish, French, Korean, English)
- ❏ Scanning process can be restarted
- ❏ Scans environment in real time
- ❏ Displays a link to an external learning page after each scan

Other Requirements

- ❏ Accurate and efficient
- ❏ Ease of use
- ❏ Free software
- ❏ Web-app optimized for mobile devices
- ❏ Camera access
- ❏ Internet connection

Architecture Overview

Client Side

Server Side   Host: Hostwinds
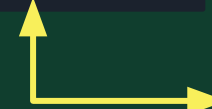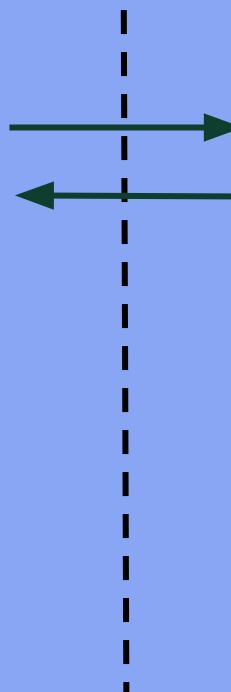

django

Object Detection App

Static Files

Text Recognition App

MVT (Model-View-Template)

Bootstrap

Collins

CSS HTML JS

Implementation Overview

Object Detection

YOLOv5

→ Python-based object detection algorithm

→ convolutional neural network

→ trainable custom object models

→ returns object classes and confidences

OpenCV

→ computer vision library

→ perform image preprocessing

→ improve the quality of input data

Figure 1: Adapted from Durve et al. (2021). The European Physical Journal Plus. 136.

Operating Principle of YOLOv5 Algorithm

Bounding Box Predictions

S X S Grid Cells

Class Probability Map

Final Prediction

Figure 1

Implementation Overview

Text Recognition

Google's Tesseract OCR

➔ pyTesseract - Python wrapper

➔ performs Optical Character Recognition (OCR) and extraction

➔ neural network with Leptonica's language-specific training data
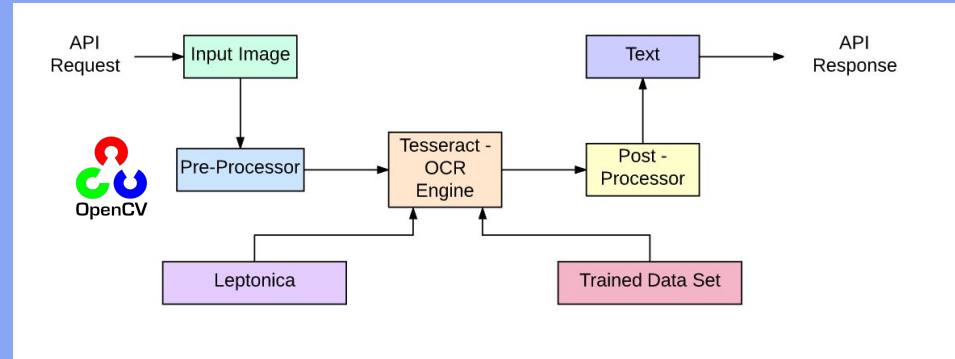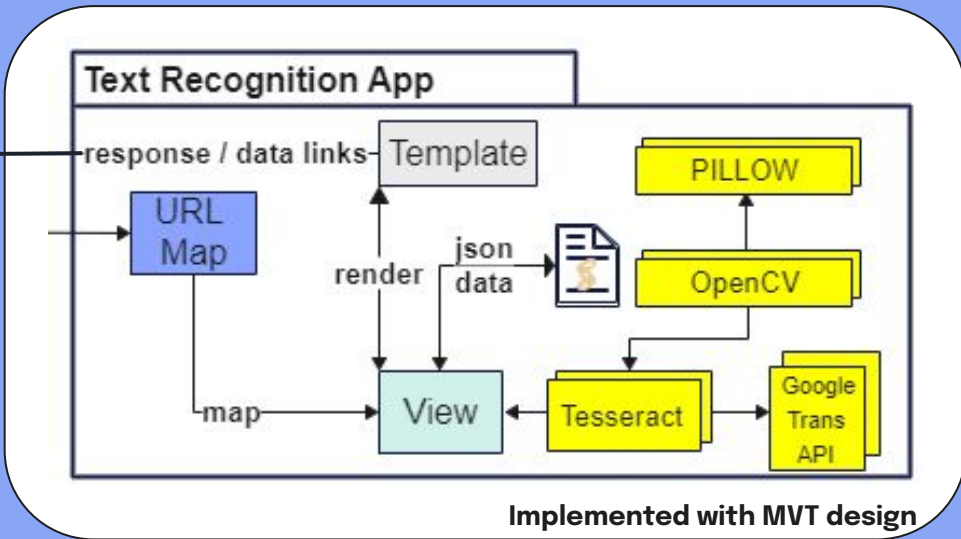
➔ collection of machine learning algorithms

Figure 2: Adapted from Parthasarathy B, (2018) Optical Character Recognition Blog
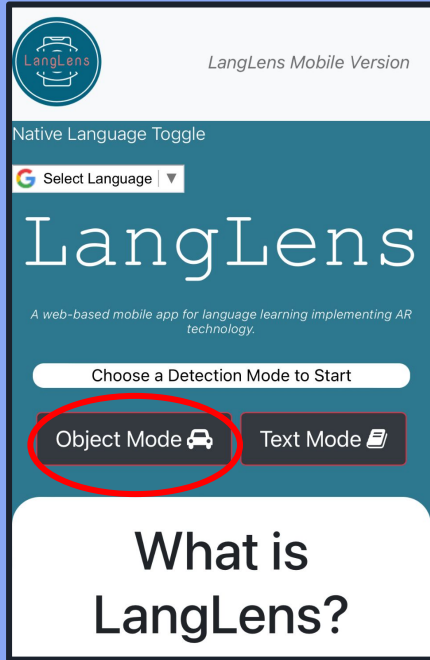
Process Flow of Tesseract OCR

Figure 2

Implementation Overview

Text Recognition Mode

Prototype Review

LangLens Mobile Version

Native Language Toggle

G Select Language ▼

LangLens

A web-based mobile app for language learning implementing AR technology.

Choose a Detection Mode to Start

Object Mode 🚗   Text Mode 📒

What is LangLens?

Homepage GUI

Click Object Mode

ℹ Guide

✕

How to use Object Detection Mode: To begin, make sure you have approved camera use for our site. Then, choose a target language you wish to translate your detected objects to. Next, aim your camera at any object in your surroundings that you wish to translate and press "Detect". *It may take a second to load on first use.* You can also press "Home" to return to the homepage, or you can toggle to Text Recognition Mode in the upper-left corner! Enjoy!

📷 Detect   🏠 Home

Guide (Object Mode)

Prototype Review

to Text Mode

Guide

✓ Spanish

English

French

Korean

Chinese

📷 Detect 🏠 Home
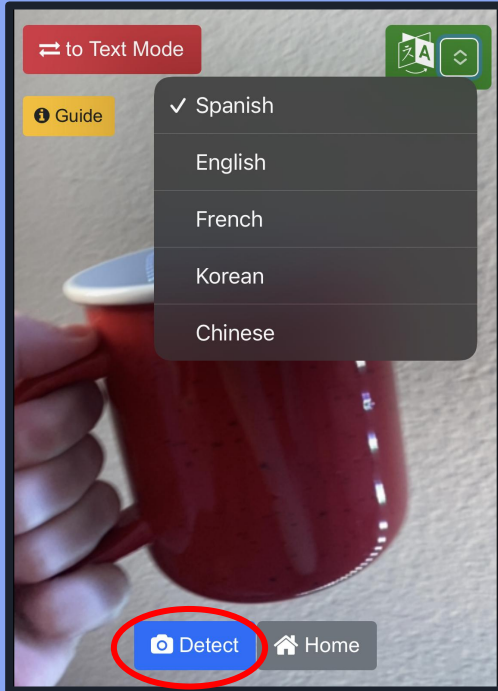
GUI (Object Mode)

Objects Detected!

Now that you've successfully detected objects in your surrounding, you can click on any of the bounding boxes to view its corresponding learning page for definitions, sentence usage, pronunciation, and more!
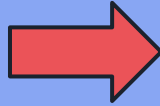
If you are unsatisfied with the resulting objects detected, hit the back button to try and scan again!

Successful Detection Popup (Objects)

Prototype Review
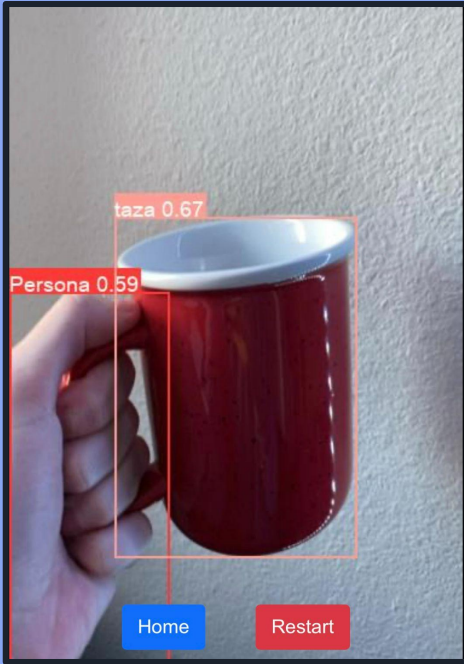
Result Image (Object Mode)

Click
bounding box

**Redirect to Learning Page
(Collins Dictionary)**

Search Query
parameter set to
" **taza** "
when redirected

Prototype Review

Result Image (Text Mode)

PASSPORT

United States of America

Home   Restart

Click
bounding box

Collins

여권

Korean   English

Dictionary   Grammar

English translation of '여권'

여권

/yeogwon/

passport

COUNTABLE NOUN

Your passport is an official document containing your name, photograph, and personal details, which you need to show when you enter or leave a country.

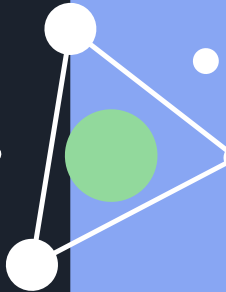You should take your passport with you when changing money. 환전할 때 여권을 가져가야 한다.

Redirect to Learning Page (Collins Dictionary)

Search Query parameter set to " 여권 " when redirected

Challenges and Resolutions

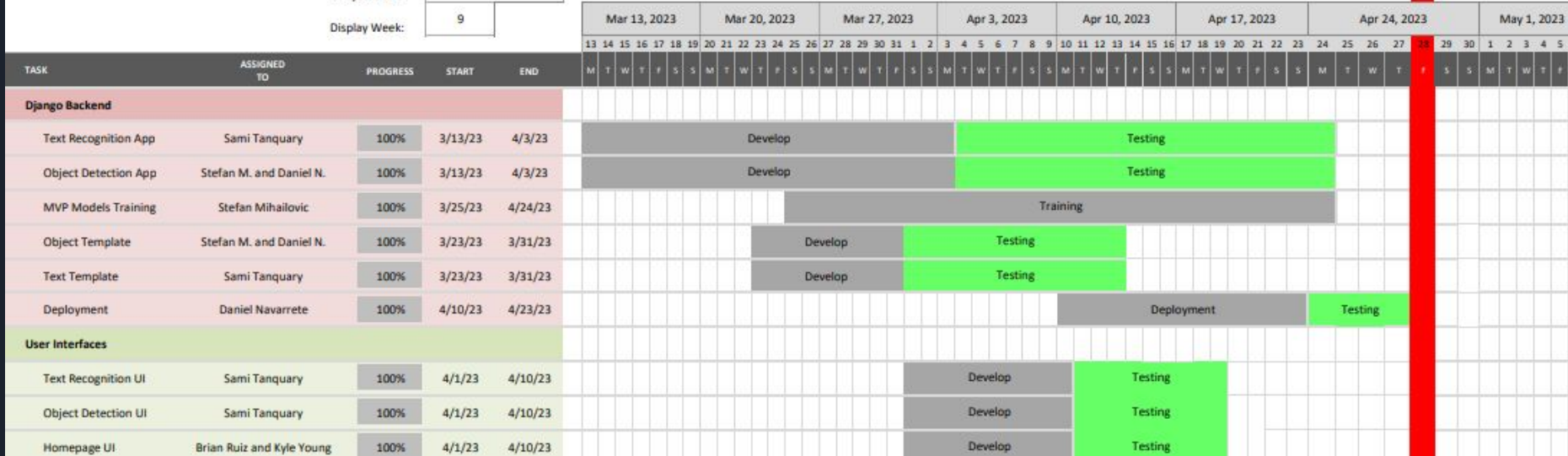| Challenge	Resolution
Low Performance running YOLOv5 / OCR in real time	Instead of running it in a real time video, scan a picture taken of the user's current environment
OpenCV can't display non-ascii characters	Convert OpenCV frame to PIL image and use imported TrueType fonts for non-ascii supported languages
Way to send user to learning page without hyperlinks	Create clickable bounding boxes around each of the objects/words that have been scanned that redirect to the external learning page.
Support as many languages as possible	Unable to support many languages for object detection mode as we need to train a language model for every desired language

Development Schedule

Development Plan

LangLens

Project Start:	Tue, 1/17/2023				
Display Week:	9				

TASK	ASSIGNED TO	PROGRESS	START	END
Django Backend				
Text Recognition App	Sami Tanquary	100%	3/13/23	4/3/23
Object Detection App	Stefan M. and Daniel N.	100%	3/13/23	4/3/23
MVP Models Training	Stefan Mihailovic	100%	3/25/23	4/24/23
Object Template	Stefan M. and Daniel N.	100%	3/23/23	3/31/23
Text Template	Sami Tanquary	100%	3/23/23	3/31/23
Deployment	Daniel Navarrete	100%	4/10/23	4/23/23
User Interfaces				
Text Recognition UI	Sami Tanquary	100%	4/1/23	4/10/23
Object Detection UI	Sami Tanquary	100%	4/1/23	4/10/23
Homepage UI	Brian Ruiz and Kyle Young	100%	4/1/23	4/10/23

= Complete

= Testing

Testing

Kind of tests	Focus	Result
Unit Tests	Tested the correct request and responses between the client and the server for the object detection and text recognition View functions.	**Found response times could be faster.** Fixed by reducing View functions to fewer steps.
Integration Tests	Tested the interactions between all system components: the camera, language, object detection, text recognition and learning page.	**Found broken pipe error if detection engines failed to detect anything.** Fixed with a fail catch that tests if the detection results are empty and show a "no text/object detected" page.
Usability Tests	Asked Users to test our application in order to get feedback from them regarding overall usability.	**Found that users wanted a loading spinner to visually show that an image is processing** and we implemented it.

Future Work

LangLens will be rebranded to *EducationalAR*

Future development on LangLens will be handled by Dr. Okim and PHD candidate, Kevin Hirschi

Integrate *internal* learning page with more learning resources

Implement *personal user accounts* to save scanned words and objects

Train the learning models with more objects for higher detection accuracy

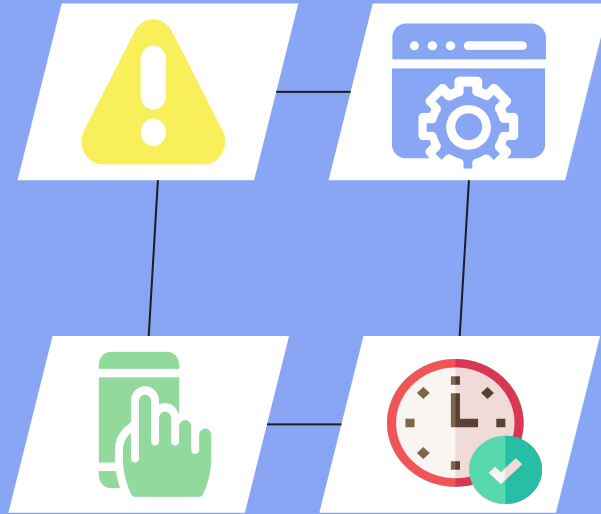Add *more languages* for learning and research (i.e. Vietnamese)

Conclusion

Problem:
- ❖ Lack of accessible AR language learning tools
- ❖ Hard for young / novice learners
- ❖ No focus on meaning, usage, and form

Solution:
- ❖ Free, easy to use, mobile web-app
- ❖ Both object / text recognition capabilities
- ❖ Learning Page with key word learning elements

Future Plans:
- ❖ Internal learning page
- ❖ User accounts
- ❖ More languages
- ❖ Train models more
- ❖ Rebrand to *EducationalAR*

Current Status:
- ❖ Completed and deployed on-time
- ❖ Dr. Okim is very happy with the final delivery