International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 11 Special Issue 02 | 2026

# Arolang: A Next-Generation Language Learning Application

Prof. Nikita Baliram Dhage[1]
[1] *Lecturer, Computer Science & Engineering, Padm. Dr. V. B. Kolte College Of Engg., M.H, India*

**DOI: 10.5281/zenodo.19254457**

**ABSTRACT**

***The goal of this project is to create an augmented reality (AR) language-learning application that allows users to interact with their environment while learning a new language. The absence of real world context in textbooks and mobile apps, among other traditional language-learning tools, may interfere and divert from the learning process. The app uses augmented reality (AR) technology to improve language acquisition by identifying objects in the user's surroundings and displaying their names in the target language. Three primary components are integrated into the system: augmented reality (AR) for visual labelling, object recognition for real-world object identification, and an API for real-time translation fetching. Users can swiftly learn the names, pronunciations, and sample sentences of items by pointing the camera at them. This practical, interactive method makes learning more engaging, improves vocabulary retention, and promotes constant practice. To help with language learning even further, the app will include features like progress monitoring and pronunciation assistance. When AR is used in this way, language learning takes on a fresh, interesting, and useful dimension that benefits learners of all skill levels and helps them expand their vocabulary in a natural way.***

***Keywords : - AR, Language Learning, Interactive, Object Recognition, Translation API, Real-Time, Visual Labeling, Vocabulary, Pronunciation, Progress Monitoring, Engaging, Practical, Skill Levels***

## 1. INTRODUCTION

Despite advances in translation technology, language barriers still hinder effective communication. Apps like Google Translate and Baidu Translate help by combining image recognition and text translation, but they struggle in situations requiring constant interaction, like driving or frequent travel. Augmented Reality (AR) offers a solution by overlaying translations directly onto the user's environment in real time, creating an immersive and hands-free learning experience. This proposed AR app uses object detection, real-time translation APIs (like Google Vision and Google Translate), and graphic overlays to label everyday objects with their names in a target language. This approach turns familiar surroundings into an interactive language-learning space, enhancing vocabulary retention through visual and contextual association. Designed for travelers, students, and those in areas with limited formal language resources, the app supports diverse users and offers potential for integration with wearable devices like smart glasses for an even more seamless experience.

## 2. . LITERATURE SURVEY

1) Paper Name: Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review Authors: Ayoub Benali Amjoud and Mustapha Amrouch
Summary: This paper reviews advancements in object detection using deep learning, CNNs, and Vision Transformers. - It classifies object detection methods into three main categories:
1. Anchor-based detectors (e.g., Faster R-CNN, SSD, RetinaNet)
2. Anchor-free detectors (e.g., YOLO, FCOS)
3. Transformer-based detectors (e.g., DETR, Swin Transformer) - Provides comparative analysis of major object detection algorithms based on speed, accuracy, and methodology.
- Discusses the evolution of object detection, from traditional methods to modern deep learning techniques.
- Covers datasets and evaluation metrics, including PASCAL VOC, MS-COCO, and ImageNet.
2) Paper Name: Real-time Object Detection Using Deep Learning
Authors: K. Vaishnavi, G. Pranay Reddy, T. Balaram Reddy,N. Ch. Srimannarayana Iyengar, Subhani Shaik (Corresponding author)
Summary:
- Traditional methods used hand-crafted features, causing slow performance and low accuracy.
- Proposes an end-to-end deep learning solution using the SSD method.
- SSD uses a single convolutional layer with depth-wise and multi-scale feature maps.

- Eliminates region proposal networks, speeding up detection.
- Achieves high accuracy, fast real-time performance, and detects multiple objects.
- Outperforms Faster R-CNN in speed with comparable accuracy.
- Future work: improve small object detection and optimize for diverse datasets.

3) Paper Name: AR Foundation for Augmented Reality in Unity
Authors: Tanisha Chaudhry, Ash Juneja, Shikha Rastogi
Summary: - Explores AR Foundation in Unity for AR development.
- Supports ARCore (Android) and ARKit (iOS) with plane detection, object tracking, and light estimation.
- Compares marker-based vs. markerless AR, highlighting markerless advantages.
- Reviews Unity's ARFoundation components like tracking, meshing, and raycasting.
- Discusses Google's ARCore, its evolution from Project Tango, and key features.
- Analyzes experimental results on plane detection, object placement, and scanning.
- Examines AR's impact on retail, gaming, training, and commercial sectors. Highlights 5G's role in improving AR's real-time rendering and interactivity.

## 3. METHODOLOGY

### 3.1 EXISTING SYSTEM

The proposed AR-based language learning app combines advanced technologies and user-centric design to deliver an immersive and effective learning experience. The development process is divided into the following key stages:

1. Object Recognition and Detection:
The application utilizes computer vision algorithms to identify objects in the user's surroundings. This process involves real time object detection using machine learning models trained on a diverse dataset to ensure accuracy across various items.

2. Augmented Reality Integration:
AR technology overlays translations and labels onto the identified objects in real time, providing users with contextual learning opportunities. The seamless blending of virtual elements with the real world enhances vocabulary retention by aligning words with tangible objects.

3. Real-Time Translation:
The app employs Google Cloud Translation API to fetch translations instantly. Captured text or object labels are processed and translated into the user's target language, ensuring a smooth and efficient translation experience.

4. User Interface Design:
A user-friendly interface is developed to simplify navigation and enhance engagement. The interface is designed to accommodate learners of all ages, featuring intuitive controls and accessible options.

5. Learning Personalization and Progress Monitoring:
The app allows users to customize their learning paths by focusing on specific topics or themes, such as household items, tools, or animals. Progress tracking features are integrated to monitor vocabulary acquisition, encouraging consistent practice and enabling users to evaluate their growth. This methodology ensures the app delivers a holistic and practical approach to language learning, making it accessible, engaging, and effective for diverse user demographics. Develop a user-friendly interface for farmers to view personalized crop recommendations, crop health monitoring, and actionable insights.

### 3.2 PROPOSED SYSTEM

This system enables real-time detection, translation, and augmented reality (AR) rendering of labels in a user-selected language for objects in the environment workflow.

**1. Mobile Device:**
- The device's camera serves as the primary input for capturing real-world objects

**2. Camera Module:**
- Captures real-time video or still images of the surroundings.
- Provides input data to the object detection module.

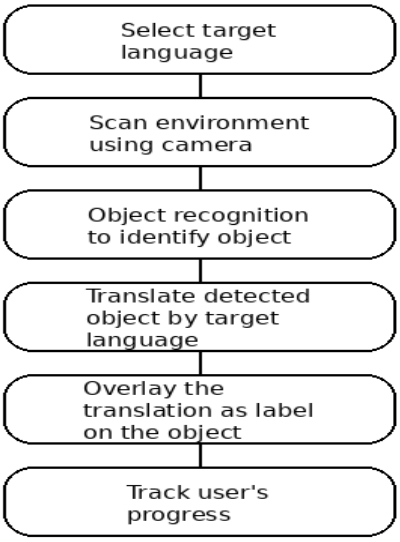
**3. Object Detection Module:**
- Utilizes machine learning models to identify objects in the camera's view.
- Outputs a list of detected objects with metadata such as object class and bounding boxes.

**4. Google Translate API:**
- Receives the names of detected objects.
- Translates the object labels into the user-selected language.

- Returns translated labels to be used in AR rendering.

**5. Overlay Module:**
- Integrates the detected and translated labels into the UI.
- Ensures that the labels are appropriately positioned relative to the detected objects in real-time.
- Enhances user experience with visually accurate and interactive overlays.

**6. Output Display:**
- Displays augmented labels overlaid on the detected objects within the mobile application.
- Provides a seamless experience for language learning and object identification.




## 4. OBJECT DETECTION

- YOLOv8n is the lightest and fastest member of the YOLOv8 family created by Ultralytics.
- Favors low latency and compact model size over detection precision perfect for lightweight deployment.
- Trained on the COCO dataset for detecting everyday objects such as bottles, people, bags, etc.
- Outputs are bounding boxes, class labels, and confidence scores per detection.

**GOOGLE TRANSLATION API**

- Google Translation API is a cloud-based service provided by Google Cloud for translating text between over 100+ languages.
- Supports real-time translation and is ideal for integrating into web, mobile, and AR/VR applications.
- Can auto-detect source language if unknown, enhancing flexibility in multilingual applications.
- Ensures scalability, reliability, and low latency, suitable for high-performance mobile applications.
- Offers client libraries for popular programming languages, including Java, Python, and Node.js.
- In this project, it is used to translate detected object labels into the user's selected language dynamically.


**Fig - Translated Label Overlay**

## 5. CONCLUSIONS

This research presents an innovative AR-based multilingual object labeling system integrating object detection and real time translation. Using AR, it ensures real-time rendering and cross-platform compatibility. The modular design allows scalability for more languages and advanced features. The system enhances language learning, accessibility, and real time translation. Future work includes improving detection accuracy, offline translation, and handling complex scenes. This system advances AR and AI in bridging language barriers for education and communication.

Clear playback controls, visual audio cues, and simplified parameter sets improve usability. Providing example outputs can help users set realistic expectations and guide interaction.

## 6. REFERENCES

[1] AR Foundation for Augmented Reality in Unity (2021)

[2] Enhancing Real-time Object Detection with YOLO Algorithm Gudala Lavanya1 and Sagar Dhanraj Pande2,(2023)

[3] Real-time Object Detection Using Deep Learning K. Vaishnavi a, G. Pranay Reddy a, T. Balaram Reddy a, N. Ch. Srimannarayana Iyengar a and Subhani Shaik(2023)

[4] Article: Long, Xiang, et al. "PP-YOLO: An effective and efficient implementation of object detector." arXiv preprint arXiv:2007.12099 (2020) .

[5] Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review(2023)

[6] Subhani Shaik, Ganesh. Taming an autonomous surface vehicle for path following and collision avoidance using deep reinforcement learning, Dickensian Journal. 2022;22(6).

[7] Z. Oufqir, A. E. Abderrahmani, et K. Satori, « Important Method for Detecting and Tracking Based on Color », vol. 8, no 5, p. 5, 2019.

[8] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175–1191).

[9] Hesamifard, E., Takabi, H., Ghasemi, M., & Wright, R. N. (2018). Privacy-preserving machine learning as a service. Proceedings on Privacy Enhancing Technologies, 2018(3), 123–142.

[10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[11] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (pp. 1097–1105). for

[12] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks large-scale arXiv:1409.1556.