

Multimodel Classification Delta Robot System

Team: Jiranjigyo

Affiliation: Department of AI Software, Korea Polytechnics Bundang Convergence Institute

Participants: Youngkyun Kim, Woohyuk Shin, Daeyoung Jung, Jinkeong Cho

Abstract

본 프로젝트는 딥러닝 기반 머신비전(YOLOv8)과 임베디드 로봇 제어 기술(OpenRB-150, AX-12 서보모터)을 결합하여, 실시간 영상 인식·분류·로봇 제어가 가능한 자동 분류 플랫폼을 개발한 사례이다. 웹캠으로 입력된 영상을 YOLO 모델이 분석하여 카드·재활용품·육류 등 다양한 객체를 실시간 분류하고, 인식 결과를 USB-TTL 시리얼 통신을 통해 로봇 제어 보드로 전송하였다. 로봇은 델타 구조를 기반으로 엔드이펙터를 제어하여 대상 객체를 지정된 위치로 정확히 이동시키며, 흡착식 그리퍼를 이용한 자동 픽업·배치가 가능하도록 구현하였다.

또한, 모바일 앱-웹 연동 UI/UX를 통해 모델 전환, 상태 모니터링, 명령 제어가 가능하고, 음성 인식(STT) 모듈을 활용해 특정 대상을 선택적(패스) 처리하는 등 사용자 상호작용 기능을 강화하였다. LCD·대시보드 등을 통한 실시간 시각화도 제공하여 학습·산업·스마트팩토리 현장에 적용할 수 있는 확장성을 입증하였다. 최종적으로 본 시스템은 단순한 객체 분류를 넘어, 교육·산업·스마트팩토리·AI·임베디드 융합교육 현장에서 활용 가능한 실질적 사례로서, 머신비전·로봇공학·임베디드 시스템을 통합한 확장형 자동 분류 플랫폼의 가능성을 제시한다.

Introduction

최근 산업 현장 및 교육 분야에서 머신비전과 로봇 자동화 시스템의 융합이 빠르게 확대되고 있다. 본 프로젝트에서는 딥러닝 기반 객체 인식(머신비전), 임베디드 제어, 로봇공학을 통합하여 트럼프 카드를 자동 분류하는 델타로봇 시스템을 개발하였다.

YOLO 기반 카드 인식, 시리얼 통신, 서보모터 제어, 음성인식 등 다양한 최신 기술을 접목하여 실시간 인식, 분류, 이동, 사용자 명령 대응까지 가능한 스마트 자동화 플랫폼의 실증 사례를 제시한다.

Methodology

Theoretical Background

머신비전 및 YOLO

YOLO(You Only Look Once)는 대표적인 딥러닝 기반 객체 탐지 알고리즘으로, 이미지 내 객체를 단일 패스(single forward pass)만으로 실시간 탐지가 가능하다는 점에서 기존 R-CNN 계열 알고리즘 대비 월등한 처리 속도를 제공한다.

본 연구에서는 최신 버전인 YOLOv8을 적용하였으며, 트럼프 카드 데이터셋을 기반으로 문양(스페이스, 하트, 다이아몬드, 클로버)과 숫자(1~K)를 정밀하게 분류할 수 있도록 전이학습(fine-tuning)을 수행하였다. 이를 통해 카드 인식의 정확도를 높이고, 실제 로봇 제어에 적합한 실시간성 객체 인식 성능을 확보하였다.

시리얼 통신 및 임베디드 제어

분류 결과는 Raspberry Pi(또는 PC)에서 OpenRB-150 제어 보드로 전송되며, 이 과정에서 USB to TTL 기반 UART 시리얼 통신(115200bps)을 활용한다.

OpenRB-150 보드는 수신된 데이터 패킷을 해석하여 실시간 모터 제어 명령으로 변환하고, 이를 기반으로 로봇 관절 제어 신호를 생성한다. 이러한 임베디드 기반 제어 구조는 통신 지연(latency)을 최소화하여 객체 인식 → 동작 수행까지의 반응성을 크게 향상시킨다.

델타 로봇 제어

델타 로봇은 3축 병렬 구조(AX-12 서보모터 기반)를 갖추고 있어, 높은 속도와 정밀도를 요구하는 픽 앤 플레이스(pick-and-place) 작업에 최적화되어 있다.

본 프로젝트에서는 순기구학 및 역기구학 알고리즘을 적용하여 좌표계 기반의 정밀한 위치 제어를 수행하였다. 이를 통해 YOLO가 탐지한 객체 위치 정보를 실시간으로 로봇 동작에 매핑시킬 수 있었다.

음성인식(STT) 기반 제어

사용자 인터페이스 향상을 위해 Google Speech-to-Text API를 적용하였다. 이를 통해 사용자가 “하트 7 선택”, “코카콜라 분류”, “플라스틱 분류”와 같은 자연어 음성 명령을 입력하면, 시스템은 이를 텍스트로 변환 후 명령 파서(command parser)를 통해 특정 분류 모델 또는 객체 동작으로 연결한다.

이로써 멀티모달 인터페이스(비전 + 음성) 기반의 지능형 제어 시스템이 구현되었다.

Implementation Method

본 연구에서는 YOLOv8 기반 멀티 분류 모델을 활용하여 트럼프카드, 재활용품, 육류, 음료 등 다양한 객체를 인식할 수 있는 구조를 설계하였다. 모바일 애플리케이션에서 사용자가 분류 대상을 선택하면, 해당 명령은 네트워크(REST API)를 통해 서버(라즈베리파이/PC)로 전송된다. 서버는 요청을 수신한 즉시 분류 모델을 교체(hot-swap)하여 선택된 객체에 대해 실시간 인식을 수행한다. 인식된 결과는 좌표 기반의 데이터 패킷으로 변환되어 OpenRB-150 제어 보드로 송신되며, 델타로봇은 해당 좌표 정보를 기반으로 자동 분류 작업을 수행한다. 또한 시스템은 LCD 인터페이스와 음성인식(STT) 기능을 포함하여 다양한 입력력 방식을 지원한다. 구체적인 절차는 다음과 같다.

- 각 분류 분야별로 YOLOv8 pt 모델을 사전 학습 및 저장
- 모바일 앱 UI에서 분류 분야를 실시간 선택
- 네트워크를 통한 명령 송수신 및 모델 핫 스왑 실행
- 카메라 영상을 실시간 인식 → 결과 패킷 전송 → 델타로봇 제어 및 자동 분류 동작 수행

Implementation Results

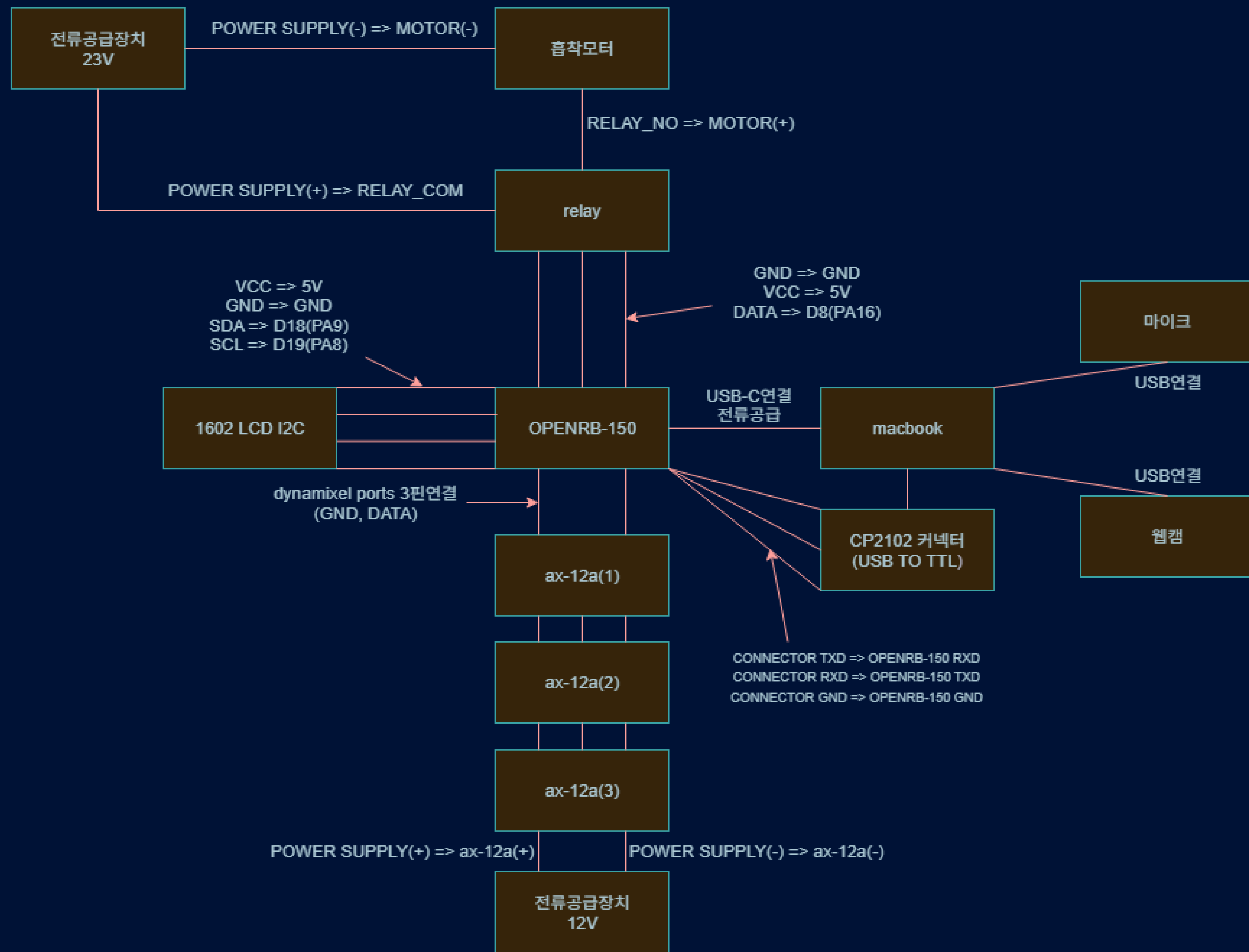
시스템은 모바일 앱에서 트럼프카드, 재활용품, 육류 등 다양한 분류 모델을 선택할 수 있으며, 사용자가 명령을 입력하면 서버는 즉시 해당 YOLO pt 모델로 전환하여 객체 인식을 수행하였다. 분류된 객체는 델타로봇을 통해 사전에 지정된 위치로 자동 이동·배치되며, LCD 화면에서는 실시간 상태 정보(프레임 속도, 연결 상태, 분류 현황 등)를 제공하였다.

또한 Google STT API 기반 음성 명령을 병행 지원하여, 사용자가 특정 대상을 제외하거나 분류 방식 변경을 지시할 수 있도록 구현하였다.

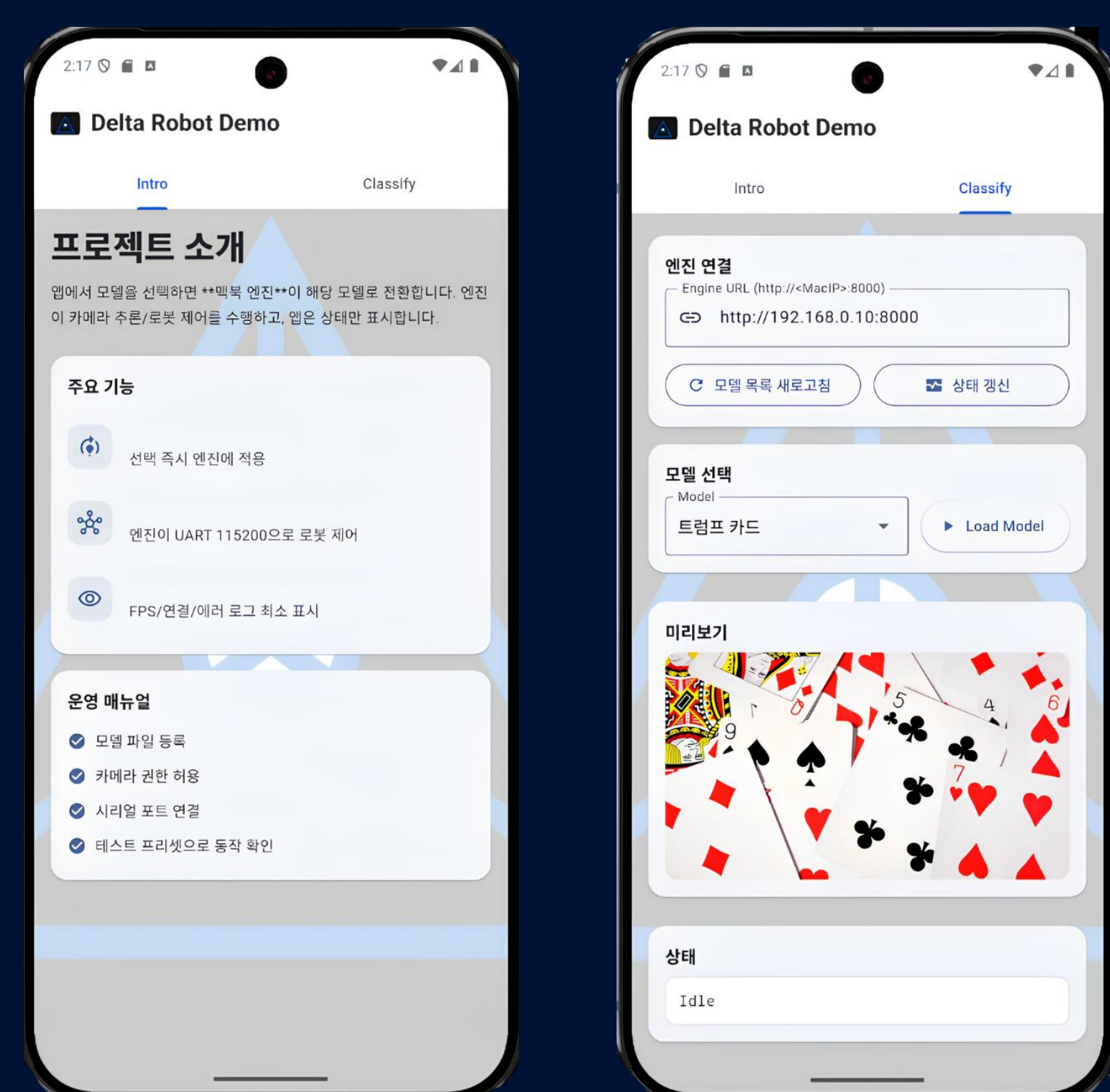
전체 시스템은 네트워크 기반 통합 환경에서 실험되었으며, 실제 분류 영상 시연을 통해 실시간성, 정확성, 상호작용성을 모두 충족함을 확인하였다.

Results

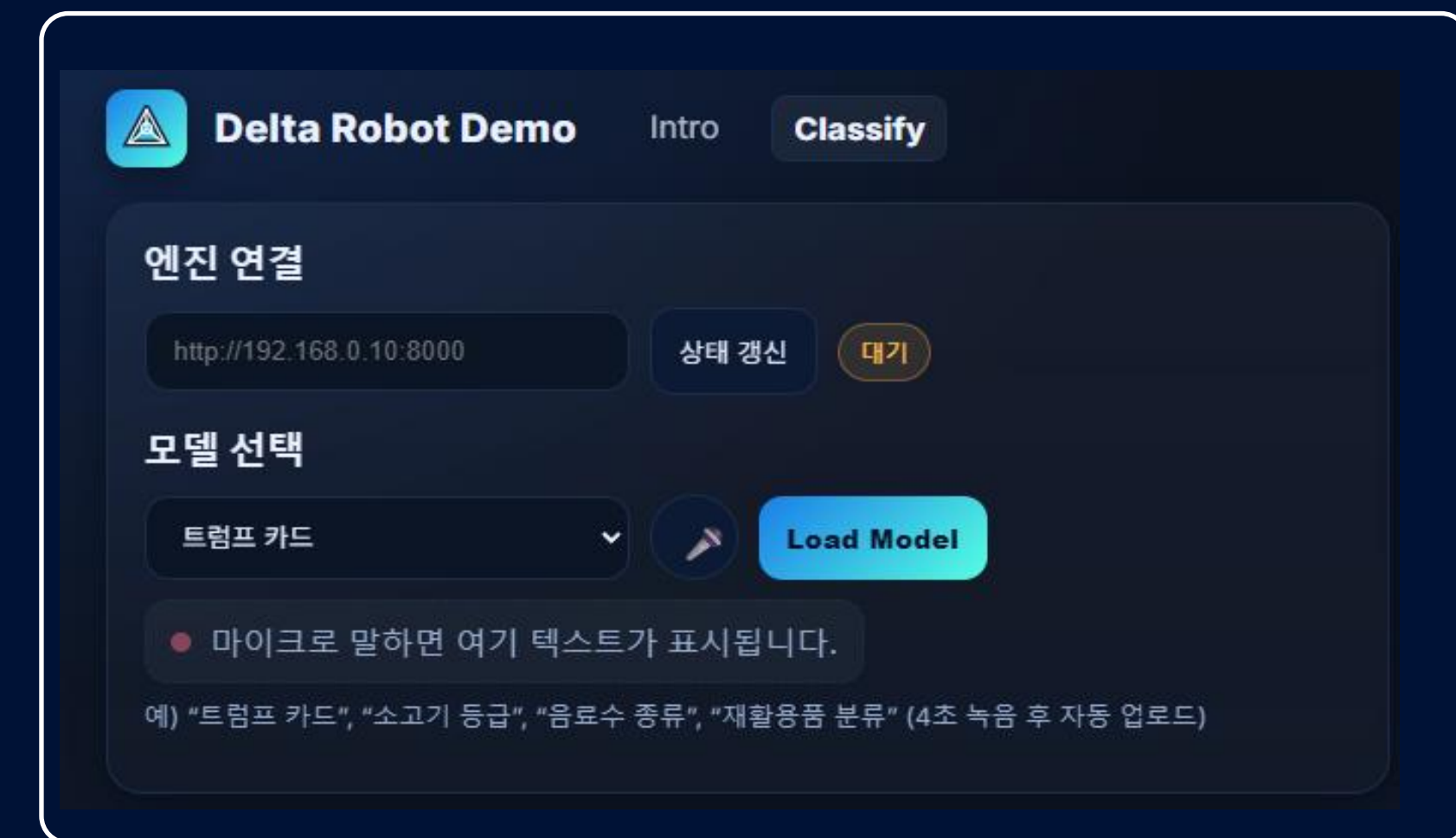
Control System Architecture



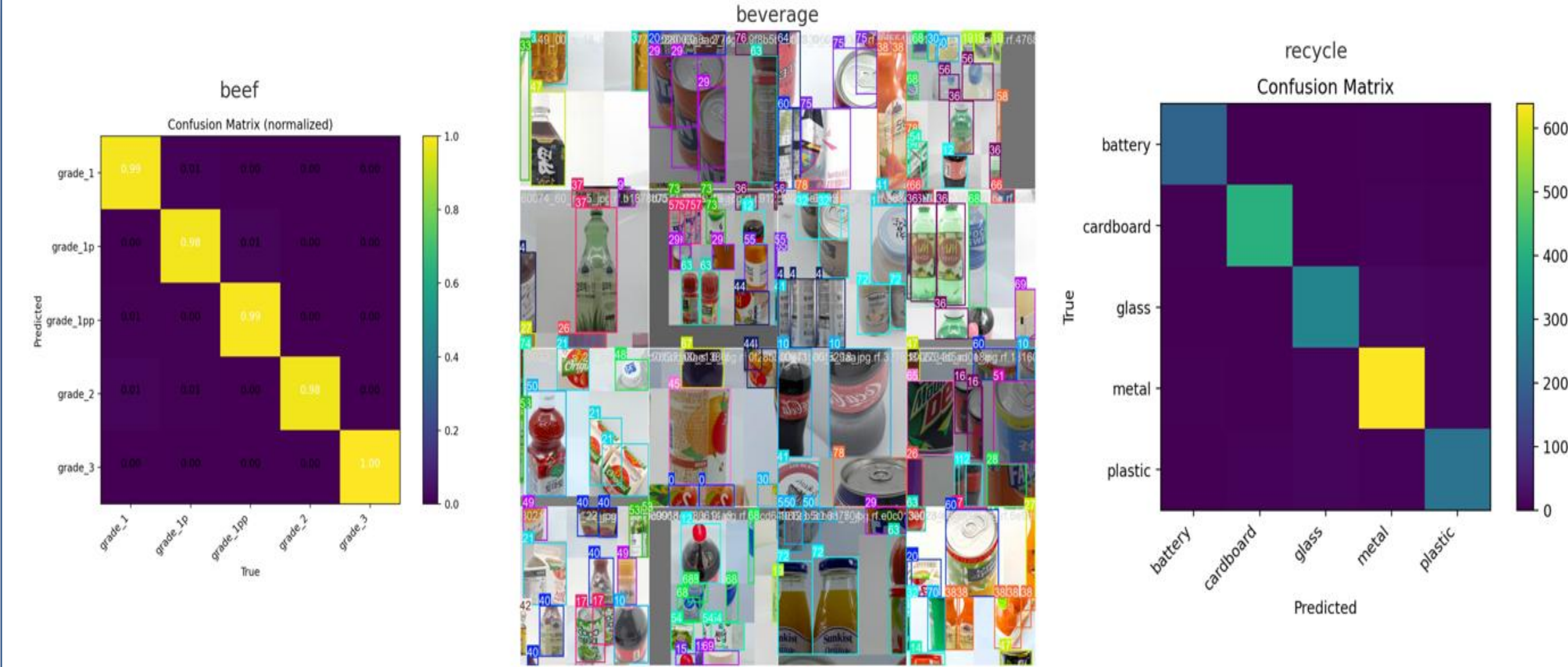
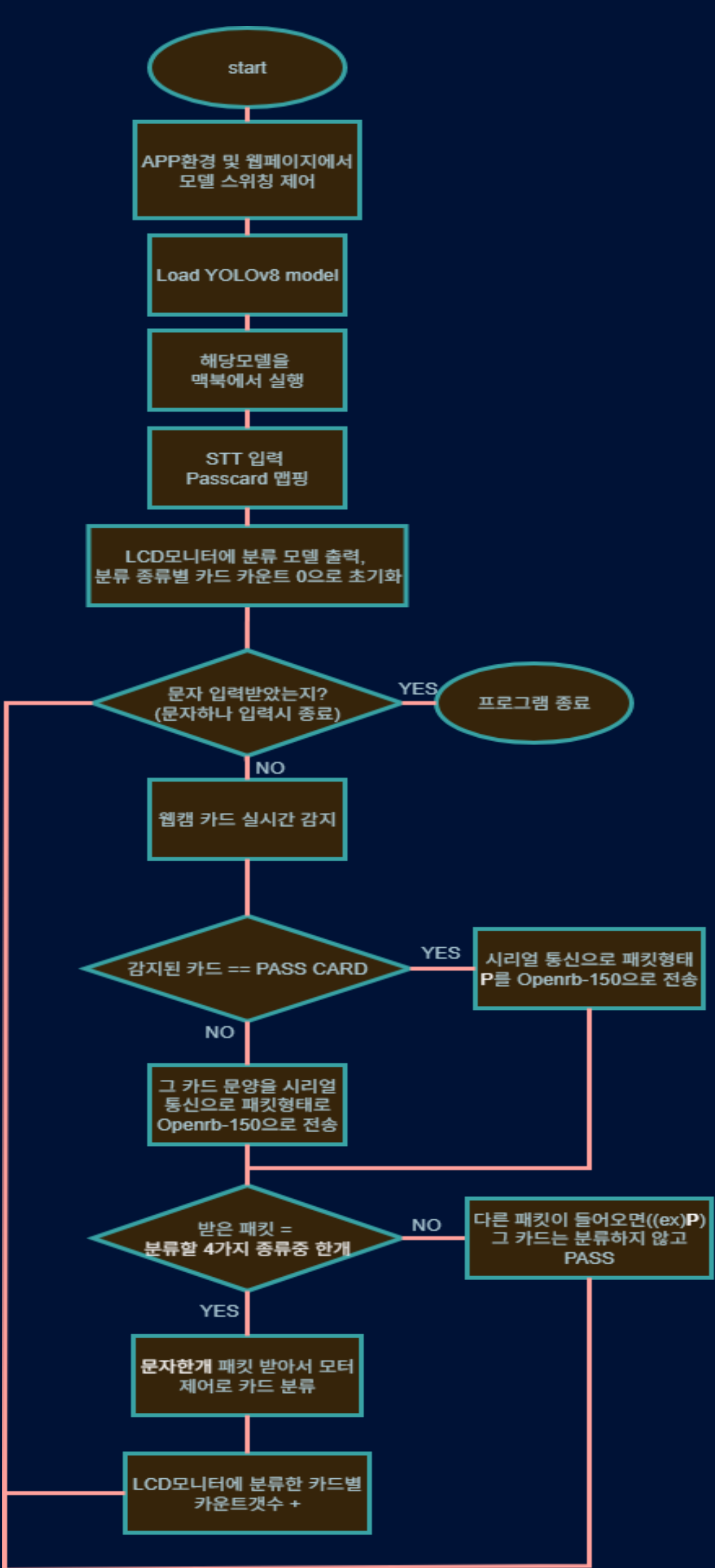
Delta Classification APP



Delta Classification Web



Flow chart



Conclusion & Discussion

본 프로젝트는 다양한 분류 모델과 사용자 맞춤 제어가 동시에 가능한 AI·로봇 융합 플랫폼을 제시한다. YOLOv8 멀티모델, 실시간 네트워크·음성제어, 직관적 모바일 UI/UX, 델타로봇 기반 분류 자동화 등 최신 기술을 통합하여, 현장 맞춤형 AI 솔루션의 실현 가능성과 활용 확장성을 입증했습니다. 현장 환경에 따라 분류 대상을 실시간으로 전환할 수 있으며, 머신비전·로봇 자동화·네트워크 통합 기술을 기반으로, 산업 현장, 스마트팩토리, 교육 등 다양한 분야에서 높은 실용성과 확장성을 가진 AI 자동화 솔루션임을 입증하였다.