

졸업 작품 개발보고서

‘비전 토크’

[시각 장애인을 위한 객체 인식 및 설명 앱]

작품명	비전 토크
과목명	융합소프트웨어종합설계1, 2
지도교수명	전병태 교수님
제출일	2024. 04. 08
팀원	김태민 (팀장)

목 차

1. 작품명

2. 작품 개요

3. 시스템 개발 필요성 및 기존 시스템과의 차별성

3.1 개발의 필요성

3.2 기존 시스템과의 차별성

3.2.1 기존 시스템

3.2.2 차별성

4. 시스템 구현 방법 및 일정

4.1 구현 방법

4.2 구현 일정

4.3 구현 분담

5. 시스템 실험

5.1 시스템 기능

5.2 시스템 기능 시험

6. 결론

6.1 시스템의 효용성 및 기대효과

6.2 개선 방향

7. 참고문헌

1. 졸업작품명

비전 토크 (VisionTalk)

2. 작품 개요

‘비전토크’ 앱은 시각 장애인과 읽기 장애를 겪는 사용자들의 일상생활 속 정보 접근성을 개선하기 위해 개발되었다. 이 앱은 사용자가 찍은 사진 속의 물체를 정확히 인식하고, 음성으로 그 물체에 대해 설명해주는 객체 인식 기능을 제공한다. 또한, 사진 속의 주요 색상을 파악하여 이를 음성으로 설명해주는 색상 인식 및 설명 기능을 통해 색상 정보를 사용자에게 제공한다. 사진에 포함된 텍스트를 음성으로 변환해주는 기능도 포함되어 있어, 시각적 정보를 오디오 형식으로 전달함으로써 시각 장애인은 물론 어린이와 읽기 장애가 있는 사람들에게 계도 큰 도움을 줄 수 있을것이다. 사용의 편리성을 높이기 위해, "비전토크"는 음성 안내와 사용자의 필요에 따라 조절 가능한 음성 응답의 속도 및 톤 조절 기능을 포함하고 있다.

3. 시스템 개발 필요성 및 기존 시스템과의 차별성

3.1 개발의 필요성

"비전토크" 애플리케이션 개발의 필요성은 현대 사회에서 정보 접근성의 중요성이 점차 증가하면서 더욱 명확해졌다. 정보가 우리 일상 생활의 모든 측면에서 중요한 역할을 하며, 시각 장애인과 읽기 장애를 겪는 사람들에게 접근 가능한 정보는 자립과 독립성을 향상시키는 핵심 요소였다. "비전토크"는 이러한 사용자들이 시각적 정보를 오디오 형식으로 쉽게 변환할 수 있도록 하여, 일상 생활에서 겪는 정보 접근의 어려움을 해결하려 했다. 이 앱은 사진 속 객체와 색상을 인식하고 설명하며, 사진에 포함된 텍스트를 음성으로 변환하여, 사용자들이 주변 세계를 더 잘 이해하고 상호작용할 수 있도록 도울것이다.

이러한 기술적 진보는 정보 접근성이 사용자의 자립성과 직접적으로 연결되어 있는 현대 사회에서 필수적일 것이다. "비전토크" 앱은 사용자들이 일상생활에서 마주하는 다양한 도전을 극복하도록 설계되었다. 예를 들어, 시각적으로 어려움을 겪는 사람들은 쇼핑을 할 때 제품의 색상을 알아야 하거나, 길을 찾을 때 방향을 알려주는 표지판의 정보가 필요했다. 이 앱은 그러한 정보를 음성으로 제공하여, 사용자가 더 독립적으로 활동할 수 있도록 지원했다.

더 나아가, "비전토크"의 개발은 팬데믹 이후 정보 접근성의 중요성이 새롭게 강조된 사회적 변화에 부응할것이다. 팬데믹 기간 동안 우리는 디지털 기술의 중요성과 AR 환경에서의

접근성의 필요성을 새삼 깨달았다. "비전토크"는 이러한 사회적 요구에 발맞추어, 시각 장애인과 읽기 장애를 겪는 사람들이 디지털 시대의 정보에 더 쉽게 접근할 수 있도록 하며, 일상 생활에서의 독립성을 향상시키는 도구가 될 것이다.

이 프로젝트는 단순히 앱 개발을 넘어서, 시각 장애인과 읽기 장애가 있는 사용자들이 사회의 완전한 구성원으로서 활동할 수 있도록 하는 데 기여할 것이다. "비전토크"를 통해, 이러한 사용자들은 자신의 능력을 최대한 발휘하고, 일상 생활의 다양한 상황에서 보다 효과적으로 정보를 접근하고 활용할 수 있게 될 것이다. 이는 디지털 접근성을 향상시키는 것뿐만 아니라, 우리 사회 전반에 걸쳐 포용성과 동등한 기회를 증진시키는 데 기여하는 중요한 단계일 것이다.

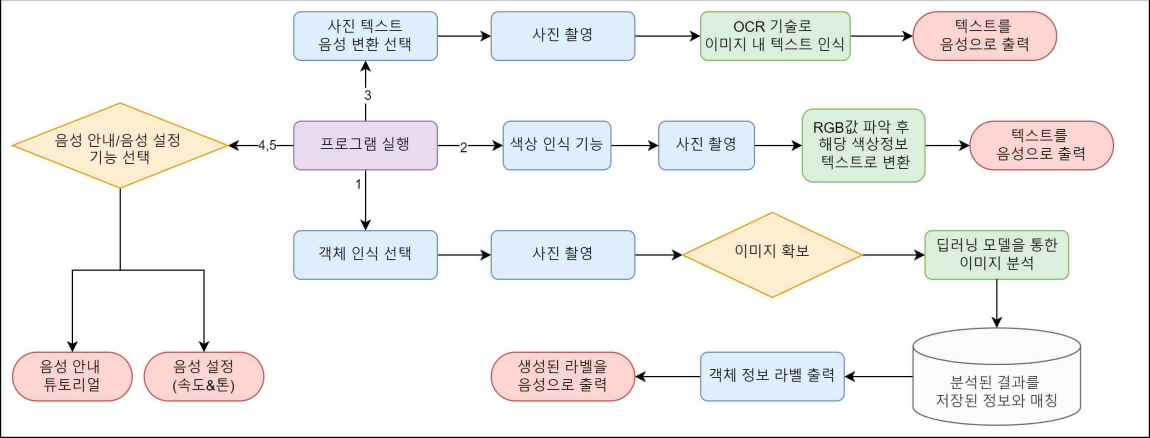


그림 1 비전토크 플로우차트 (후일 첨삭)

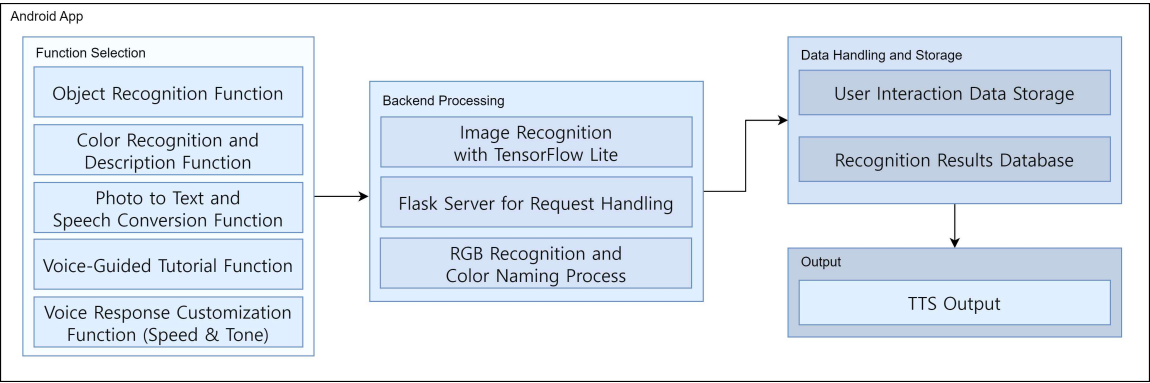
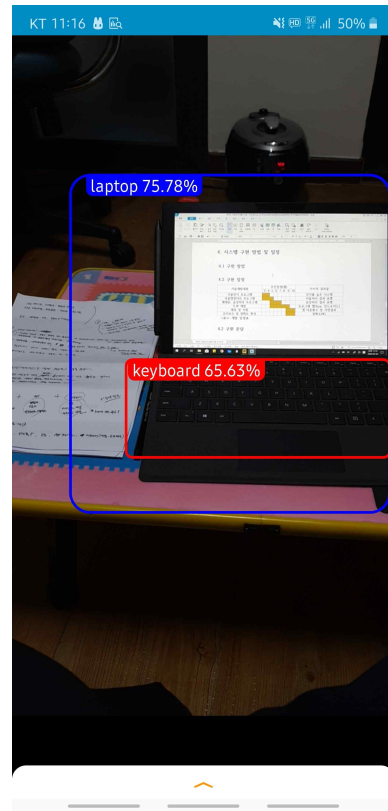
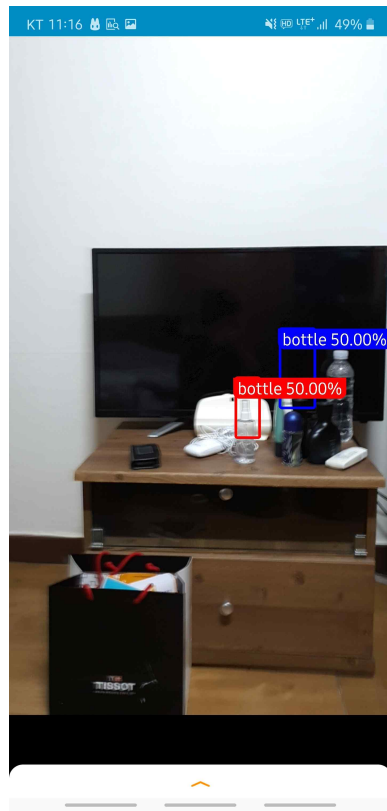


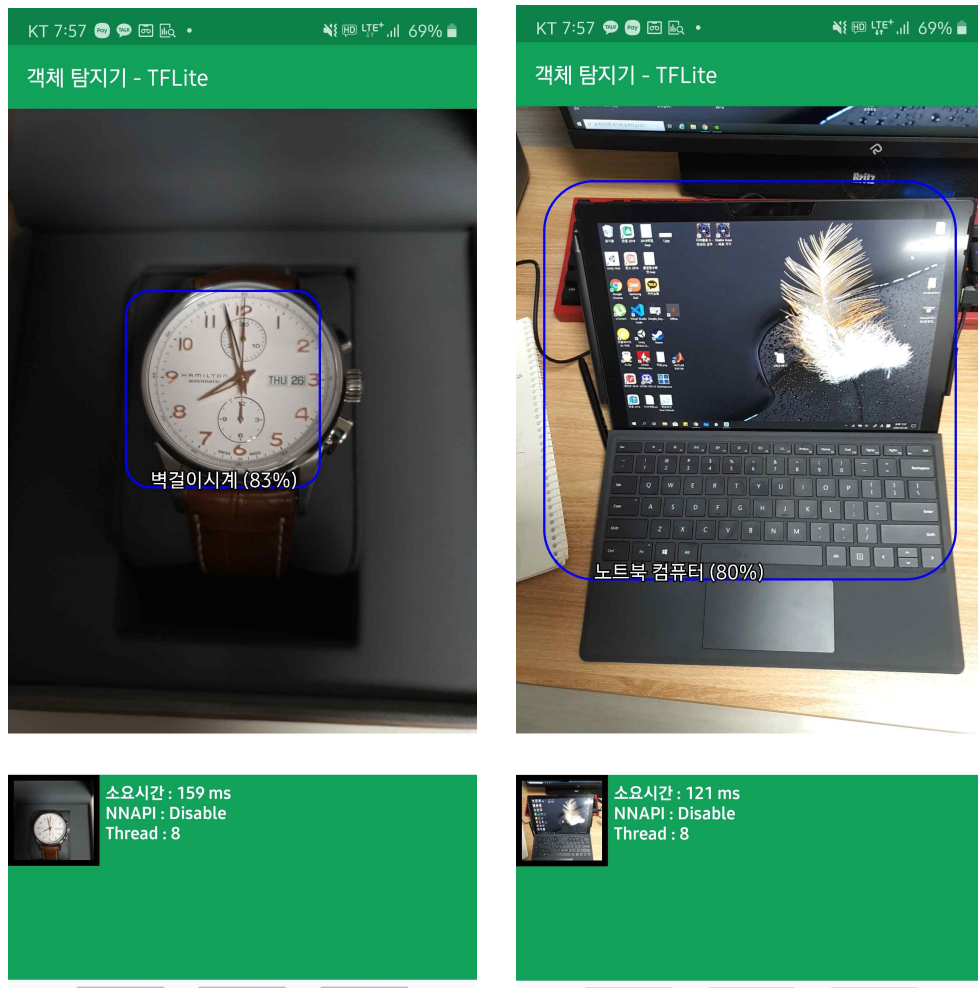
그림 2 비전토크 시스템 순서도

3.2 기존 시스템과의 차별성(후일 작성)

3.2.1 기존 시스템(후일 작성)



<그림1> Object_detection 안드로이드 앱 사용 화면

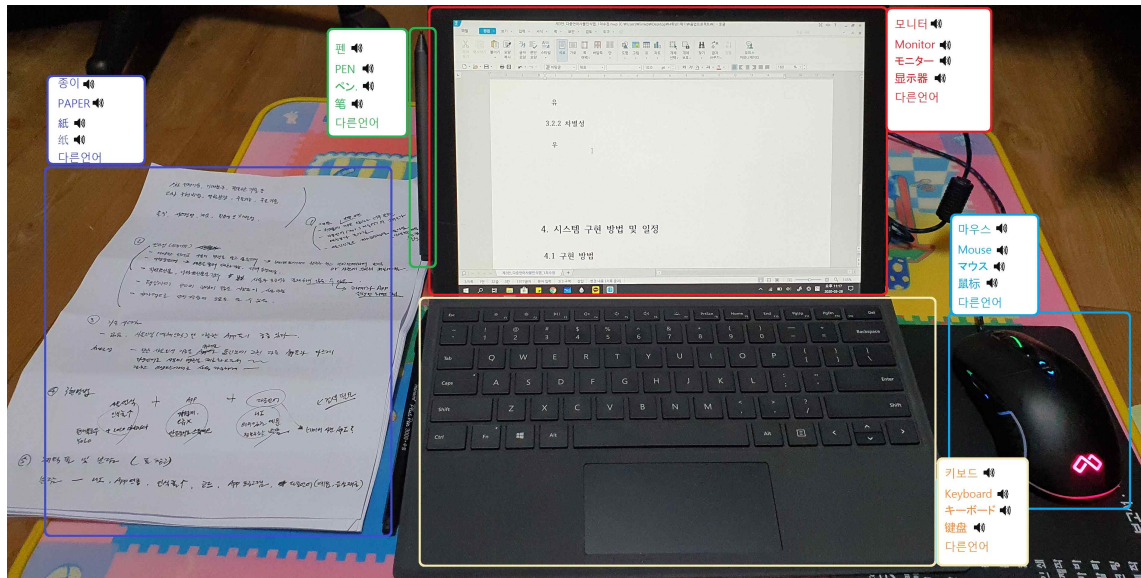


<그림2> TFLite-객체탐지기 앱 사용 화면

<그림1>의 Object_detection 은 아주 기본적인 텐서플로우 객체인식 프로그램을 사용한 앱이다. 다중언어로 사물명칭을 실시간 지원하지는 않으나 설정에서 원하는 언어를 선택하는 것은 가능하다. 그리고 사물의 인식 속도는 빠르나 객체를 분간하는 인식률이 저조하고 dataset의 정보가 적어 인식된 객체의 명칭을 제대로 출력하지 못한다.

<그림2>의 TFLite-객체탐지기는 국내 개발자가 프로토타입으로 제작한 앱이다. NNAPI를 사용한다고 하지만 프로토타입이다보니 신경망 프레임 워크 자체를 아직 지원하지 않는다. 인식률은 이전 앱보다 좋은 편이고 한글 명칭만 지원한다.

3.2.2 차별성(후일 작성)



<그림3> 본 프로젝트 개발 앱의 사용 예상도 1 (프로토타입 완성되면 이미지 변경필요)



<그림4> 본 프로젝트 개발 앱의 사용 예상도 2(프로토타입 완성된 후 이미지 변경)

한국 고유의 사물들을 실시간 인식하여 외국인 근로자들에게 그들의 언어로 사물의 쓰임새, 발전 배경을 설명하는 앱이다.

4. 시스템 구현 방법 및 일정

4.1 구현 방법

VisionTalk 앱은 TensorFlow Lite와 Keras를 사용하여 시각 장애인을 위한 객체 및 색상 인식 기능을 구현한다. 주요 기능은 3가지 이고 주요 기능을 보조하기 위해 2가지 기능이 있다. 첫째로 객체 인식 기능이다. 객체 인식 기능은 tensorflow lite 와 keras를 이용해 시각 장애인을 위한 객체 인식 기능을 가진 CNN 딥러닝 아키텍처를 구현한다. 사용자가 촬영한 이미지는 딥러닝 모델을 통해 분석되고, 객체에 대한 정보를 반환한다. 앱은 반환된 정보를 TTS 기능을 통해 음성으로 변환한다. 둘째로 색상 인식 기능이다. 색상 인식 기능은 사용자가 촬영한 이미지의 대다수를 차지하는 색상의 RGB값을 파악한 뒤, 이를 색상 정보를 텍스트로 변환 후 TTS 기능으로 음성으로 출력한다. 셋째로 사진 촬영 시 OCR 기술로 텍스트를 인식하고 추출한 후, TTS 기능을 통해 이를 음성으로 출력한다. 보조를 위한 2가지 기능은 주요 기능을 안내해주는 음성 안내 기능이고, 다른 하나는 TTS기능의 음성의 속도나 톤을 조절해주는 음성 설정 기능이다.

4.2 구현 일정

개발 내용	추진일정(월)									가시적 결과물
	3	4	5	6	7	8	9	10		
아이디어 회의										주제 선정 및 보고서 작성
개발 환경										개발 환경 선정 및 환경 구축
처리 및 학습 모델 개발										이미지 처리 및 모델 훈련
계획 회의										개선사항 추가 및 수정
앱 인터페이스 개발										인터페이스 개발 및 기능구현
통합 및 시스템 테스트										평가 및 개선사항 토의
부족한 점 보완										보완 및 마무리작업

기간		일정	내용
3월	3.15 ~ 3.31	아이디어 회의	프로젝트 개발 주제 선정, 보고서 작성
4월	4.01 ~ 4.24	개발 환경	개발 환경 선정 및 환경 구축
5월	5.01 ~ 5.31	이미지 처리, 기계 학습 모델 개발	이미지 처리 파이프라인 구축 및 CNN 모델 훈련
6월 ~8월	6.01 ~ 6.07 7.01 ~ 7.07	계획 회의	개선사항 추가 및 수정, 추가 기능 개발에 대한 세부 계획 수립
	5.01 ~ 8.31	앱 인터페이스 개발 및 기능 구현	인터페이스 개발과 동시에 기능 구현
9월	9.01 ~ 9.30	통합 및 시스템 테스트	모든 기능과 인터페이스 통합, 종합적인 시스템 테스트 및 버그 수정
10월	10.01 ~ 10.31	부족한 점 보완	부족한 부분 보완 및 마무리 작업

4.3 구현 분담

2010280010_김태민

- 기획, 보고서작성, 개발환경 및 DB 구축, 이미지 처리 및 기계 모델 학습 개발, 앱 인터페이스 개발, 통합 및 시스템 테스트

5. 시스템 실험

5.1 시스템 기능

5.1.1 객체 인식 및 음성 설명

5.1.2 색상 인식 및 음성 설명

5.1.3 이미지 내 텍스트 음성 변환

5.1.4 바코드 인식 및 음성 설명

5.1.5 음성 응답의 속도나 톤 조절

5.2 시스템 기능 시험(시연) (후일 작성)

6. 결론 (후일 작성)

6.1 시스템의 효용성 및 기대효과

6.2 개선 방향

7. 참고문헌 (후일 작성)

모델 평가와 성능 지표 확인 진행

주간 보고서 1

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 4월 1일

업무구분	내 용
완 료	알약 이미지 수집, 개발 방향성 및 보고서 내용 수정
진행중	이미지 크기 조정, 밝기 조절, 배경 제거등 전처리 작업
예 정	모델 훈련 및 모델 평가, API 요청 구현을 위한 DOCKER 컨테이너 이용해 텐서플로우 서버 준비. 보고서 개선 할 점 없나 보기 (시스템 구조도).

주간 보고서 2

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 4월 8일

업무구분	내 용
완 료	이미지 크기 조정, 밝기 조절, 배경 제거등 전처리 작업 모델 훈련 및 모델 평가, API 요청 구현을 위한 DOCKER 컨테이너 이용해 텐서플로우 서버 준비.
진행중	<p>주제 탐색 물체 인식 기능 사용 1. 시각 장애인을 위한 객체 인식 및 설명 앱(저시력자 시각장애) 2. 독서 보조 앱 (촬영한 텍스트 음성 변환 기능, 시각 장애 or 어린이들 or 읽기장애가 있는 분을 위한) 3. 식물 or 동물 or 식물과 동물 인식 앱 (교육적인 방향) 물체 인식 기능 사용 X 1. 소외계층을 위한 교육 및 직업 훈련 플랫폼 2. 농촌 지역 디지털 격차 해소 앱 3. 비콘 기반 도움 요청 앱</p> <p>객체 인식 앱 할시 추가기능(예상) 1. 객체 인식 및 설명 기능(객체를 tts로 음성으로 말해줌) 2. 색상 인식 및 설명 기능(색상을 tts로 음성으로 말해줌)(근데 시각 장애가 있을 시에 이 색이 무슨 색상인지 인식 자체를 못 할텐데 해당 기능이 의미가 있는가?, 후천적 시각 장애를 위해?) 3. 앱 사용법 음성 안내(설명서...) 4. 음성 응답의 속도나 톤 조절 기능(옵션기능, 시각장애가 있는 분이 어떻게 설정을 바꿀것인가? 핸드폰 왼쪽의 음량조절바를 이용?) 5. 텍스트 음성 변환 기능(주제탐색 2랑 겹칠 듯, 차라리 주제탐색 2의 기능을 객체인식 앱 하위로 넣는게 좋을지도) 추가? : 질감?texture? 근데 텍스처가 의미가 있나? 그냥 만져보면 되는데? 날씨? 돈인식(이게 얼마짜리 동전 or 지폐인가,근데 이건 손의 촉감이랑 지폐의 길이로 판별한다고 한다, 또 요새는 삼성페이나 카드쓴다. 여기에다가 예전에 만들던 의약품 문답 추천 기능 넣으면 괜찮을 듯, 근데 비전 토크랑 문답추천이 무슨 연관인가? 텍스트 음성 변환 기능에 아래의 독서보조앱 요약 요청 기능 넣어도 괜찮을 듯 바코드를 통해 물건 인식 후 tts로 말해주는 기능)</p> <p>독서 보조 앱 1. 촬영한 텍스트를 번역해주는 시스템(이미 갤럭시 기본 내장 기능임) 2. 읽기속도조절, 다양한 음성 선택 3. 사전 참조 기능. 4. 요약기능(텍스트를 ai 모델에 전송해서 요약을 요청하는 식, 해당 모델 호스팅 서버와 API통신 필요, openai에서 제공하는 ai모델 chat-gpt와의 api를 openai에서 제공해줌)</p> <p>앱 이름은? 1:객체인식 2:독서보조 2인포소닉(info+sonic),1비전토크(vision+talk),1뷰토크(view+talk),2인사이트가이드,1아이에코(eye+echo)</p> <p>객체 인식 앱을 기반으로 보고서 작성</p>
예 정	보고서 개선 및 기능 개발

주간 보고서 3

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 4월 15일

업무구분	내 용
완 료	주제 탐색
진행중	kaggle에서 COCO 모델 YOLOv5 tensorflow lite 모델 확보, flask를 이용해 간이 서버 제작, 테스트 이미지를 이용해서 객체 인식 기능 정상 작동 확인.
예 정	앱 아이콘 만들고 기능 만들기

주간 보고서 4

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 4월 22일

업무구분	내 용
완 료	kaggle에서 COCO 모델 YOLOv5 tensorflow lite 모델 확보, flask를 이용해 간이 서버 제작, 테스트 이미지를 이용해서 객체 인식 기능 정상 작동 확인.
진행중	앱 스플래시 아이콘 제작 및 객체 인식 아이콘 제작 앱 스플래시 화면과 메인 화면 제작
예 정	TTS 기능 및 여타 기능 제작

주간 보고서 5

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 4월 29일

업무구분	내 용
완 료	앱 스플래시 아이콘 제작 및 객체 인식 아이콘 제작 앱 스플래시 화면과 메인 화면 제작
진행중	음성 응답의 속도나 톤 조절 기능 구현 OCR TTS 기능 제작 (kor.traineddata, eng.traineddata)
예 정	시스템 기능 중 다른 하나

주간 보고서 6

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 5월 6일

업무구분	내 용
완 료	음성 응답의 속도나 톤 조절 기능 구현
진행중	OCR TTS 기능 제작중(Google ML Kit로 변경)
예 정	색상 인식 기능 제작

주간 보고서 7

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 5월 13일

업무구분	내 용
완 료	
진행중	
예 정	

주간 보고서 8

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 5월 20일

업무구분	내 용
완 료	
진행중	
예 정	

주간 보고서 9

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 5월 27일

업무구분	내 용
완 료	
진행중	
예 정	

주간 보고서 10

(교수님께 보고용으로 임시 제작, 졸업작품 개발보고서 최종
제출시 주간 보고서 전체 삭제 예정)

■ 보고일자 : 2024년 6월 3일

업무구분	내 용
완 료	
진행중	
예 정	