2023 제1회 K-디지털 플랫폼 AI 경진대회

멀티모달을 활용한 시각장애인 보조 애플리케이션 ViewFinder

팀명 엔드포인트

팀장 조세은

팀원 김가람, 양정열, 이승준

CONTENTS

팀원 구성

조세은

총괄 및 개발 지원

양정열

AI 모델링

김가람

전략 기획 및 서버 구축

이승준

애플리케이션 개발 및 서버 구축

01

개발 배경 및 방향

사회적 현황

도입 배경

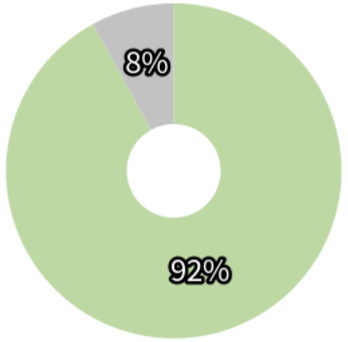
ViewFinder

사회적 현황

시각장애인의 디지털 격차

세계 디지털전환 시장 개요 [1]

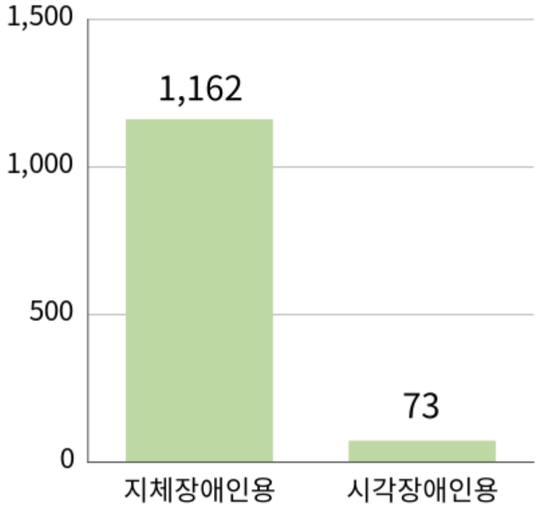
현재(2021년) 규모	
4844억 4000만 달러(약 629조 7720억원)	
2021~2030 연평균 성장률 전망	**14.9%**
2030년 시장 규모 전망	
1조 6924억 달러(약 2190조 6430억원)	
최대 시장	**미국**
가장 빠르게 성장 중인 시장	**아시아태평양**



[시각장애인의 모바일 앱 사용 설문조사]

2022년 한국소비자원 설문조사에 따르면, 쇼핑 및 모바일 앱 사용 경험이 있는 시각장애인 중 약 92%가 사용상 어려움을 겪었다고 응답함 [2]

시각장애인 보조기기 지원 미비

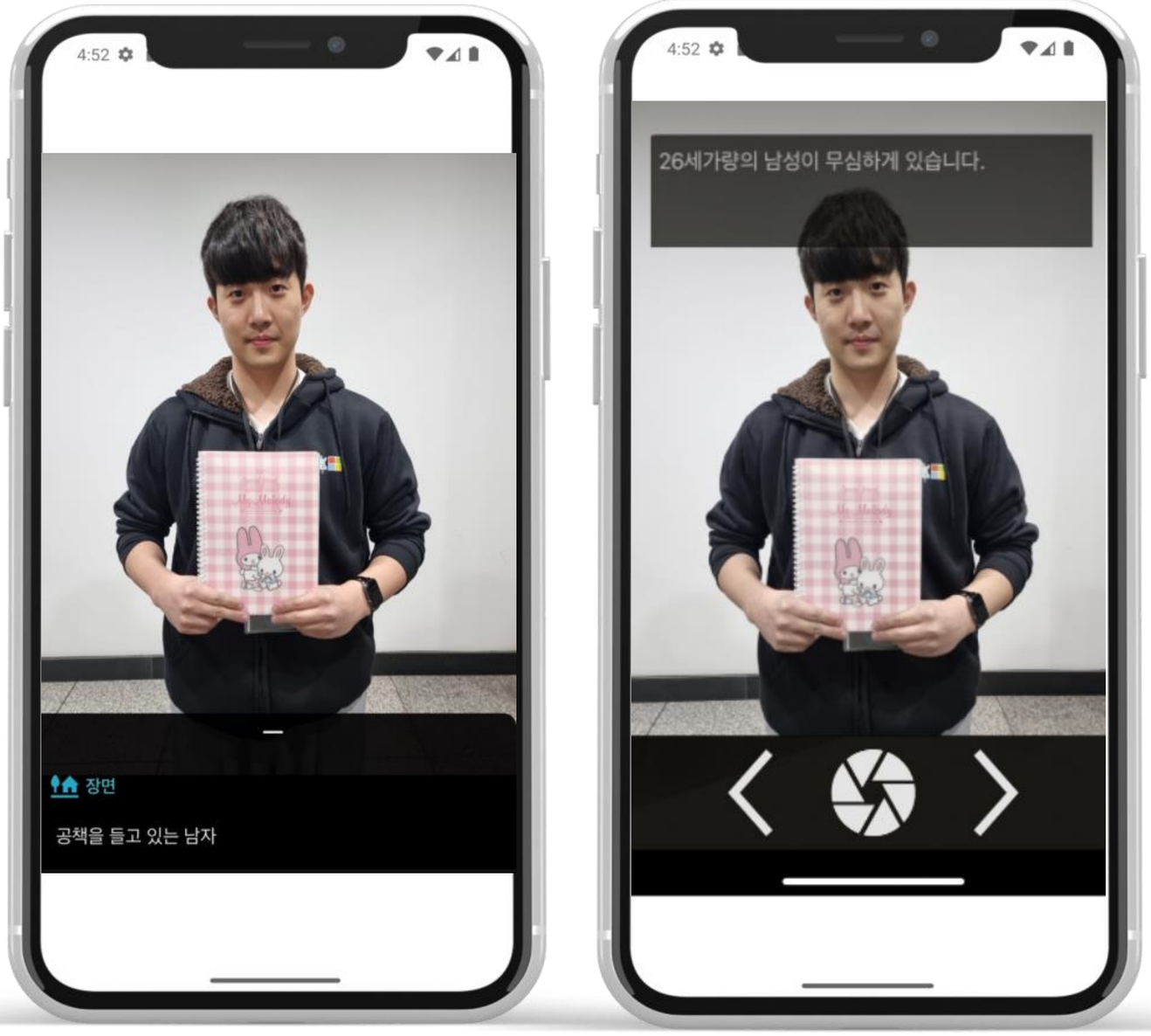


서울시 장애인용 보조기기 보유 현황

서울시 4개 센터의 지체 장애인용 보조 기기는 1,162개인 반면, 시각장애인용 기기는 73개에 불과함 [3]

일상이 디지털화 되고 있으나 시각장애인들은 디지털의 이점을 충분히 누리기 어려움

도입 배경



Seeing AI

설리번 +

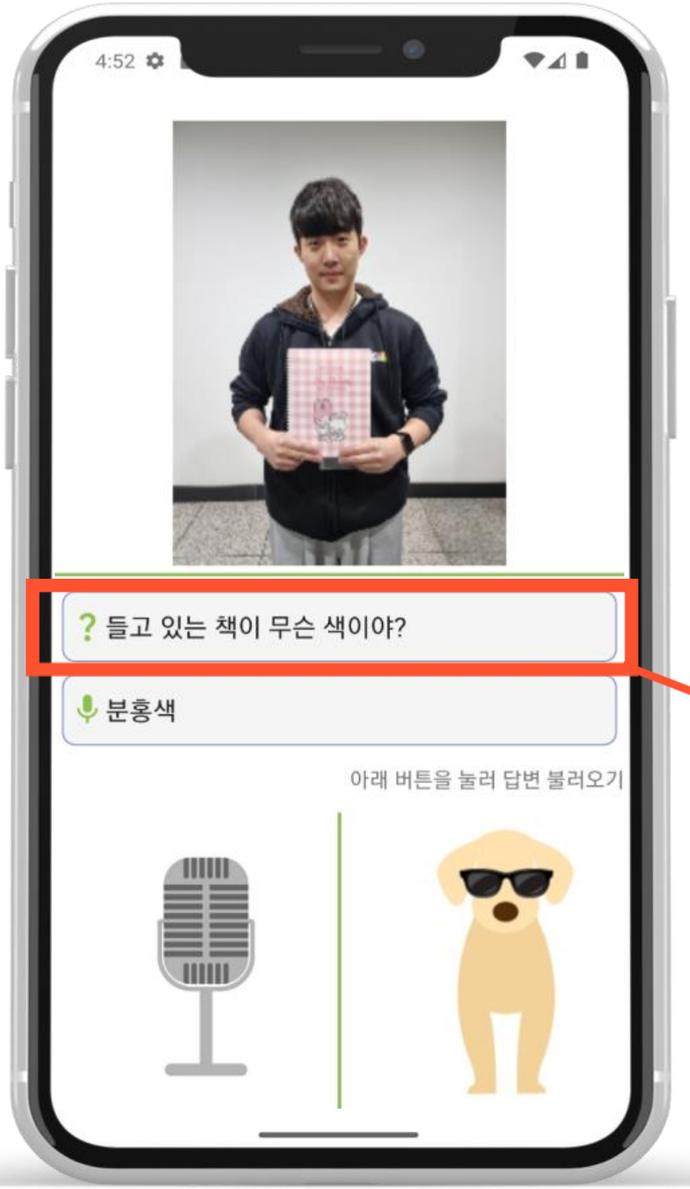
Captioning

기존 애플리케이션은 주어진 상황에 대해 정해진 표현으로 설명

하지만, 사용자가 원하는 정보는 다를 수 있다.

ex) 사람이 들고 있는 공책은 무슨색일까?

ViewFinder



ViewFinder

- 시각장애인들에게 자세한 정보를 제공하기 위한 애플리케이션 ViewFinder를 제안함

KEY POINT!!

Questioning 기능을 도입함으로써 사용자가 원하는 정보를 얻도록 함

02

개발 프로세스

데이터 수집
데이터 전처리
모델 아키텍쳐
하이퍼파라미터 세팅
애플리케이션 아키텍쳐

데이터 수집

활용 데이터 셋

- 인간의 상식이나 배경지식을 바탕으로, 이미지 관련 질문에 대해 이미지 속에서 답을 찾아야 하는 Task

AI Hub ..... 외부 지식 기반 멀티모달 질의응답 데이터 [4]

Image 수 : 60,084
질문 & 답변 수 : 120,168

Example 1

Q) 이미지 속 피자는 몇 판이야?

A) 1

Example 2

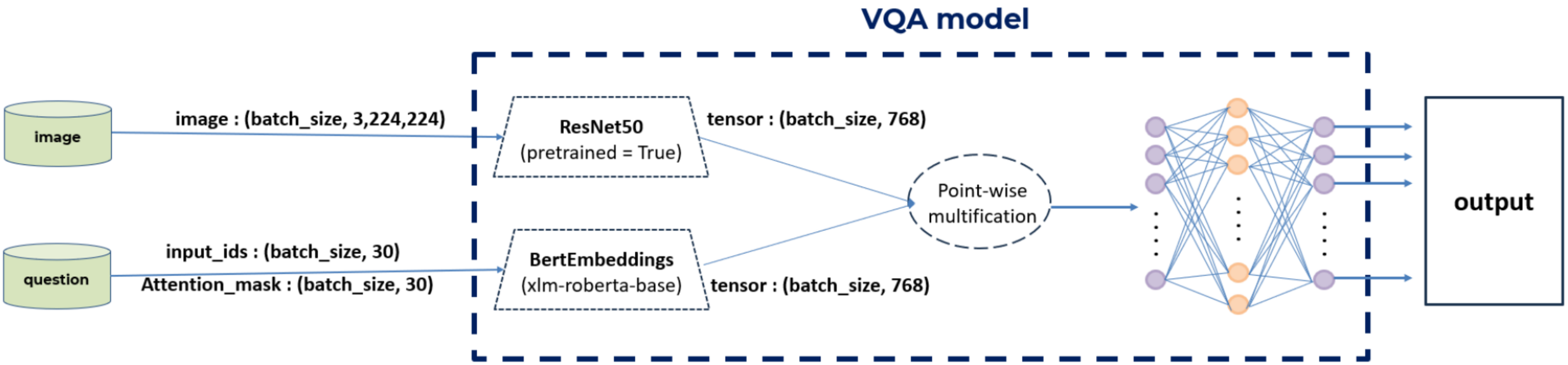
Q) 오른쪽에 있는게 뭐야?

A) 자동차

데이터 전처리

	전처리 방법		
데이터	Input	Image	Torchvision.transforms 패키지로 전처리 Resize(356,356), RandomCrop(224,224), Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225] 전처리 후 tensor 크기 -> (batch_size,3,224,224)
		Question	Hugging face의 'xlm-Roberta-base' tokenizer로 input_ids, attention_mask 추출 add_special_tokens = True, max_length = 30, truncation = True, pad_to_max_length = True 전처리 후 tensor 크기 -> input_ids : (batch_size, 30) -> attention_mask : (batch_size, 30)
	Output	Answer	하나의 대답이 하나의 인덱스를 매칭하도록 딕셔너리를 구축 가장 많이 답변된 상위 1,000개의 답변을 활용

모델 아키텍쳐



Pretrained ResNet50 : 14,000,000개의 이미지를 1,000개의, 클래스로 분류해 사물의 특징(색, 모양 등)을 이해한 모델 [5]
XLM-RoBERTa : 100개 언어가 포함된 2.5TB의 필터링된 CommonCrawl 데이터로 사전 학습된 모델 [6]

하이퍼파라미터 세팅

❖Model 설정

Resnet	
Pretrained	True
fully_connected_dim	768

XLM-RoBERTa	
position_embedding_type	absolute
vocab_size	250002

VQA Mdel	
hidden_dim	1054
output_node	1000

❖ Hyperparameter 설정

- Loss_function = CrossEntropyLoss
- Optimizer = AdamW (lr = 0.00002)
- Batch_size = 30
- epoch = 30

❖ 성능평가

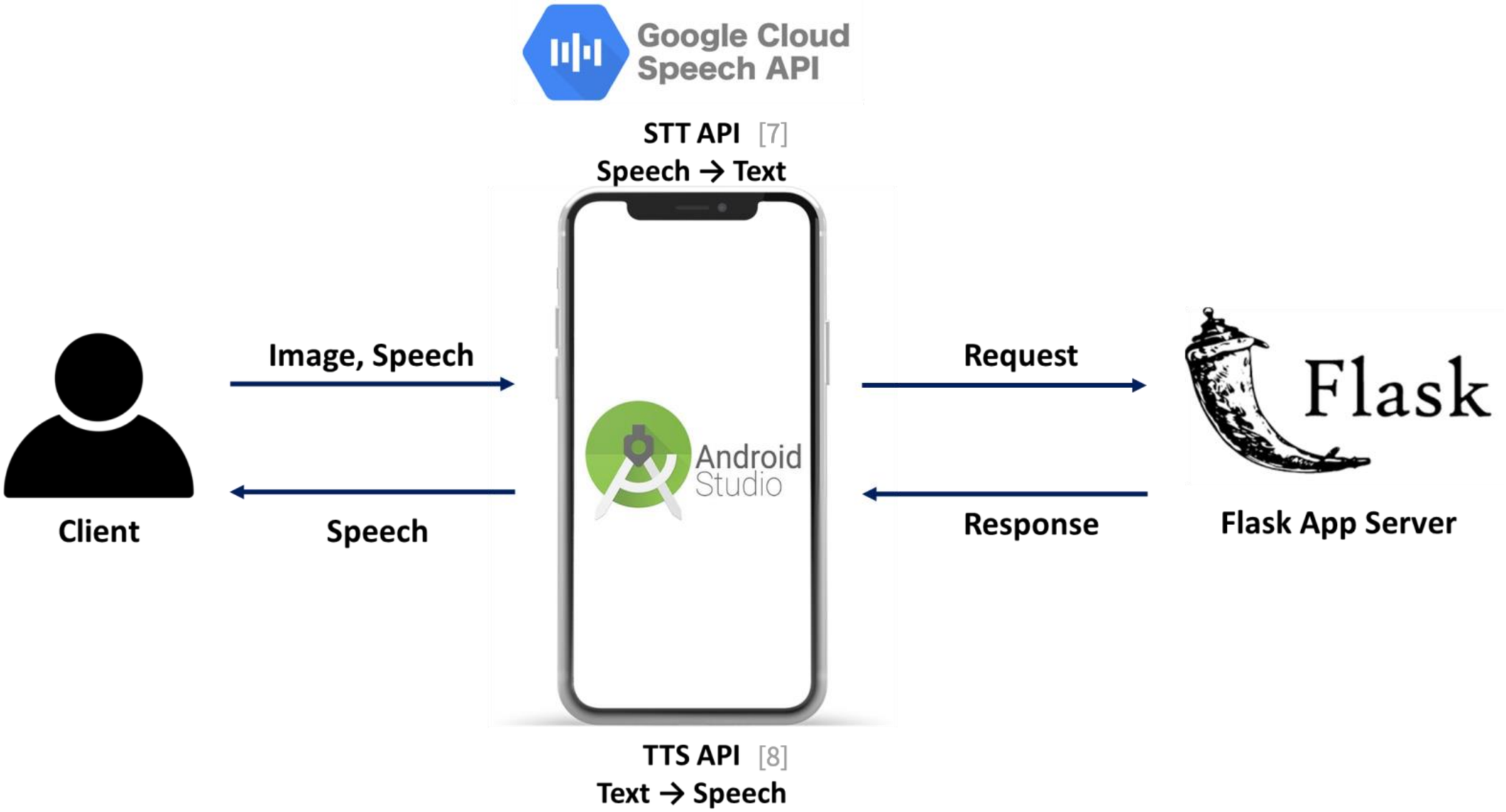
Best_accuracy = 0.8944

Loss = 0.00325

❖학습 및 테스트 데이터 셋 구성

- 데이터셋 120,168
- train : test = 96,134 : 24,034

애플리케이션 아키텍쳐


Google Cloud Speech API
STT API [7]
Speech → Text
Image, Speech
Request
Android Studio
Flask
Client
Speech
Response
Flask App Server
TTS API [8]
Text → Speech

03

시제품 형태 및 활용 방안

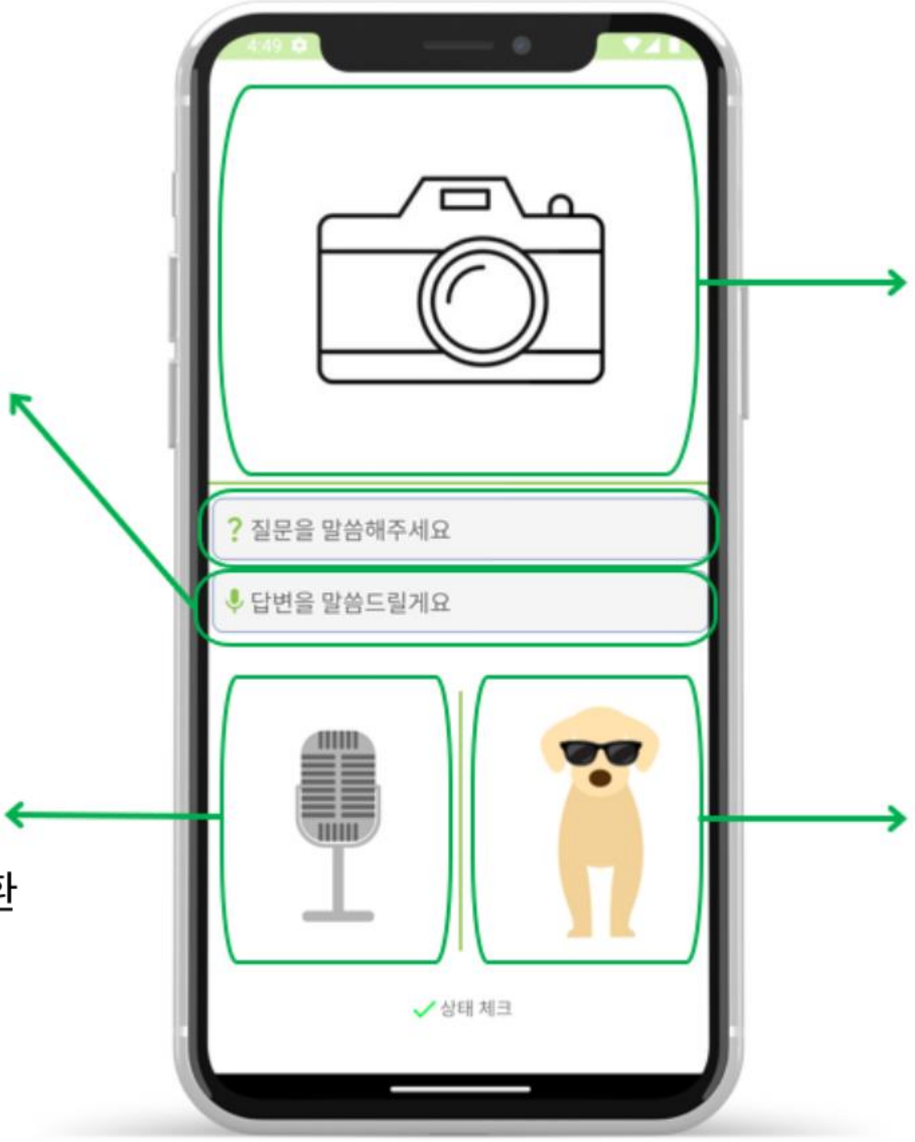
시제품 형태

ViewFinder 시연 영상

시제품 형태

④ **텍스트 음성 변환 Text-To-Speech**
서버로부터 받아온 결과값을 텍스트로
나타내고 음성으로 변환하여 출력

② **음성 인식 기능 Speech-To-Text**
궁금한 점을 질문한 내용이 텍스트로 변환



① **사진 촬영 기능**
확인하고 싶은 물체 or 상황 촬영

③ **Request & Respond**
입력한 이미지와 질문을
서버에 보내고 결과값을 반환

ViewFinder 시연 영상

https://www.youtube.com/shorts/B6P652DpLWk

04

기대효과

기대효과

- **디지털 혁신의 사각지대에 있는 사람들에게 배리어프리 환경 제공**

  시각, 청각 등 다양한 장애를 가진 사람들도
  디지털 기술을 활용하여 정보의 접근과 소통이 가능

- **시각장애인들의 이동권 보장**

  사진에 대한 구체적인 설명을 제공함으로써
  시각장애인들이 주변 환경을 더 정확하게 인식

- **복지 산업 및 기타 산업으로의 기술 확장 가능**

  시각장애인과 색맹인에게 도움을 줄 수 있는
  보조 장치로서 복지 산업에 활용

참고문헌

[1] "2030년 시장 규모 2000조원… 디지털전환 격전지로 뜬 한국." 서울신문, 2022. 12. 05, https://www.seoul.co.kr/news/newsView.php?id=20221206017013

[2] 한국소비자원, 장애인 소비자 모바일 거래 실태조사 보고서

[3] “서울시 장애인보조기기센터, 시청각장애인용 기기 부족” 에이블뉴스, 2022.11.24, https://www.ablenews.co.kr/news/articleView.html?idxno=100882

[4] https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=71357

[5] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[6] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[7] https://developer.android.com/reference/android/speech/SpeechRecognizer

[8] https://developer.android.com/reference/android/speech/tts/TextToSpeech

'AI로 더 나은 미래'

2023 K-디지털 플랫폼 AI 경진대회

THANK YOU

엔드포인트