presentation

멀티모달 AI 기반 포켓몬 이미지 및 캡션 분석 모델 개발

윤성헌

presentation_index

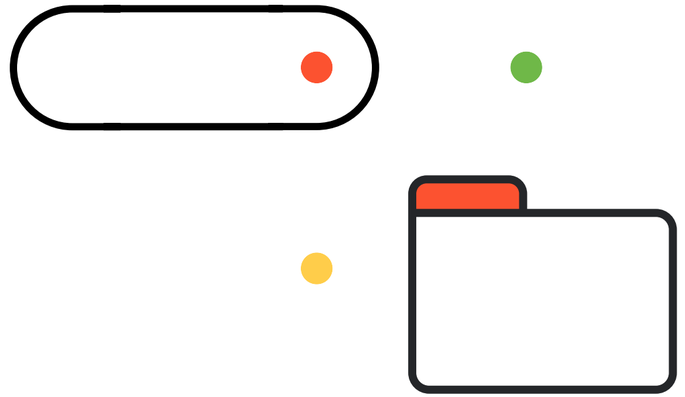
INDEX

목차

01

인트로 (Introduction)

1. 프로젝트 배경 및 문제 정의
2. 프로젝트 목표
3. 기대효과



프로젝트 배경 및 문제 정의

- 기존의 이미지 분류는 주로 이미지 자체의 정보만 사용해서 진행되는 경우가 많음
- 하지만 현실은 이미지와 함께 텍스트 설명이 같이 존재하는 경우가 많음
- 이미지만 단독으로 처리하는 모델보다 이미지와 텍스트를 동시에 이해하고 처리하는 멀티모달 (Multimodal) AI 모델이 대세
- 그렇기에 이미지와 캡션 데이터를 함께 학습하는 모델을 개발하는 필요성 느낌

프로젝트 목표

- 포켓몬 이미지와 그에 따른 캡션(설명 문장)을 같이 분석하는 멀티모달 AI 모델을 만드는 것이 목표
- 이미지 특징 추출에는 사전 학습된 EfficientNet 모델을 활용
- 텍스트는 문장 토큰화를 통해 LSTM 신경망으로 처리해 두 정보를 결합

기대 효과

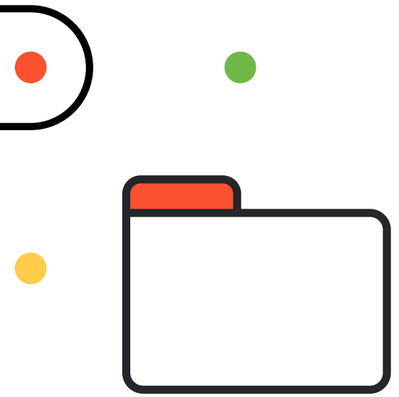
- 이 모델이 성공하면 이미지와 텍스트 사이의 관계 이해도를 높임
- 자동으로 이미지에 태그를 붙이거나, 원하는 이미지를 더 정확히 찾는 검색 시스템 개발에 활용
- 전자상거래나 SNS 같은 다양한 서비스에서 콘텐츠를 더 똑똑하게 분류하고 추천하는 데 도움

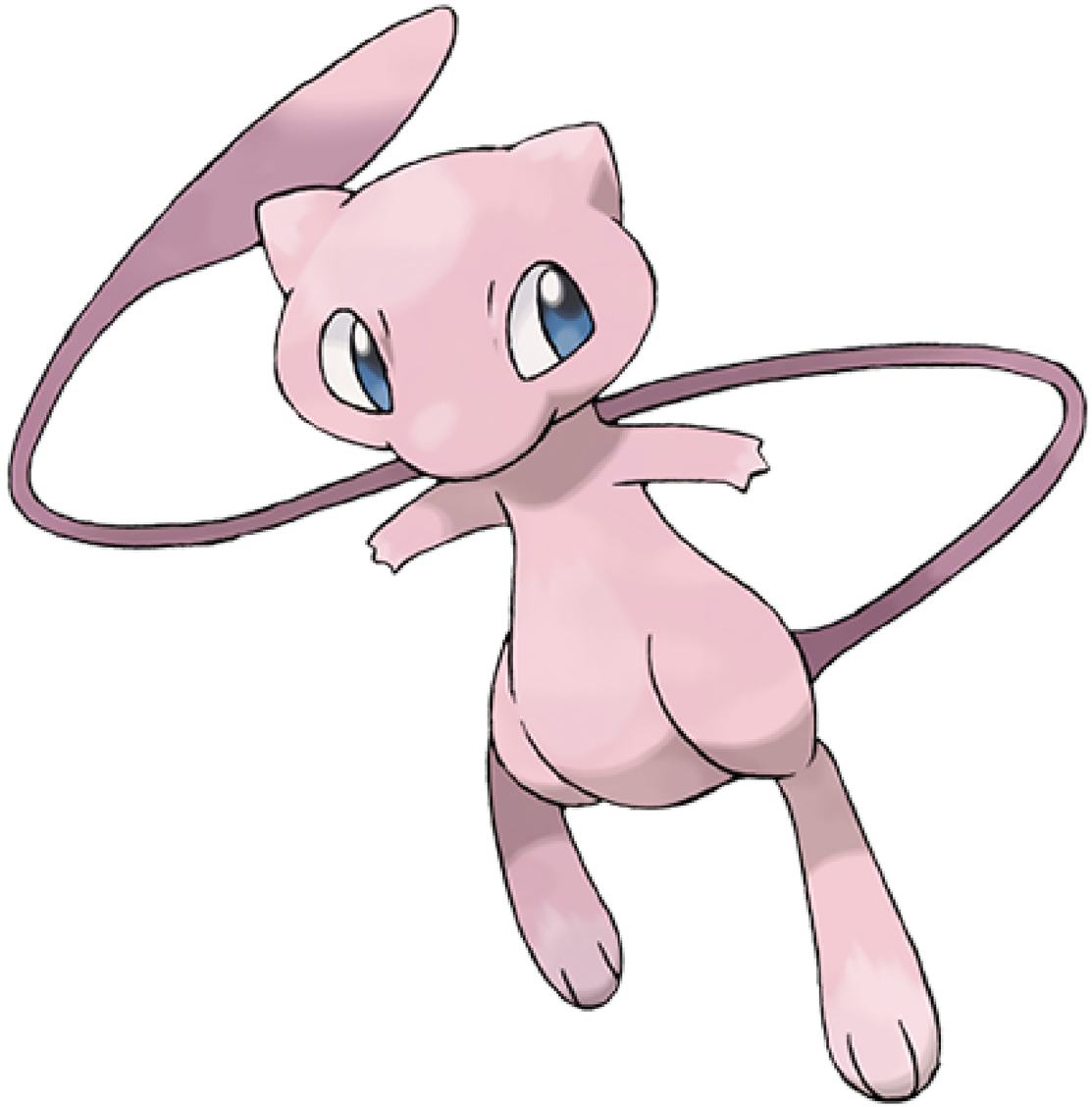
02

데이터 준비 (Data Preparation)

1. 데이터셋 소개

2. 데이터 전처리(Preprocessing)





presentation_photo

데이터셋 소개

- 포켓몬 이미지와 해당 이미지의 설명 텍스트(캡션) 사용
- 이미지와 캡션 데이터 불러와 실제 존재하는 이미지와 유효한 캡션만 선별
- 모든 이미지는 224×224 크기로 리사이징
- 수천 개 이상의 이미지-캡션 쌍으로 구성
- 다양한 포켓몬 캐릭터와 특징 포함

데이터 전처리 (Preprocessing)

특징 융합 및 분류

- 이미지와 텍스트에서 나온 특징 연결.
- 이어서 Dense 레이어를 여러 개 거쳐 최종 sigmoid 활성화 함수로 이진 분류 결과 출력

이미지 브렌치

- 224 X 224 RGB 이미지
- 사전학습된 EfficientNetB0 모델을 사용해 이미지 특징을 추출
- 추출된 특징은 GlobalAveragePooling2D를 통해 벡터화되고, BatchNormalization과 Dropout으로 정규화 및 과적합 방지 처리

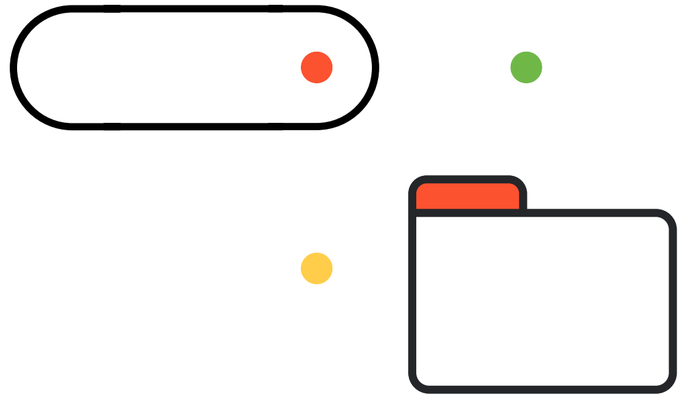
제목은 문장보다 키워드

- 입력: 토큰화 및 패딩된 캡션 시퀀스 (최대 길이 고정)
- Embedding 층으로 단어를 128차원 벡터 공간에 임베딩
- LSTM 층으로 문장의 순차적 특성과 의미를 파악
- 이후 BatchNormalization과 Dropout을 거쳐 특징 벡터 생성

03

모델링 (Modeling)

1. 전체 아키텍처
2. 핵심 기술 및 모델 선정 이유
3. 학습 환경



전체 아키텍처

이미지 입력 → EfficientNetB0 → GAP → BN → Dropout

\

→ Concatenate → Dense → BN → Dropout → Output

/

텍스트 입력 → Embedding → LSTM → BN → Dropout

핵심 기술 및 모델 선정 이유

EfficientNetB0

- 적은 계산량으로도 높은 정확도를 달성
- ImageNet 등 대규모 데이터셋에서 우수한 성능 입증
- 사전학습 가중치 활용으로 빠른 수렴과 좋은 특징 추출 가능

LSTM

- 순차 데이터(문장)의 순서와 관계를 잘 반영하는 RNN 계열 모델
- 긴 문맥 정보도 기억해 자연어 처리에 효과적
- 캡션과 같은 텍스트 시퀀스의 의미를 파악하는 데 적합

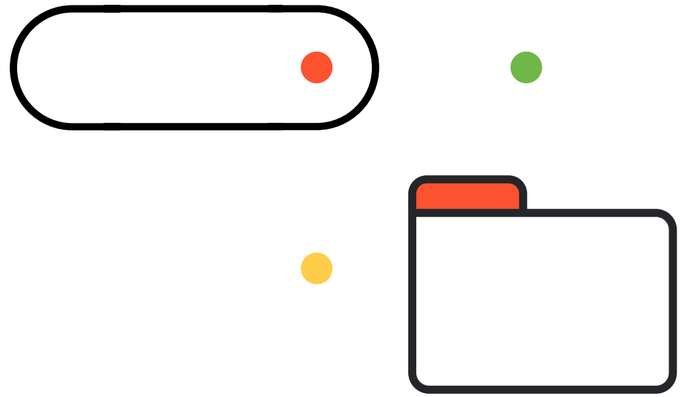
학습 환경

- 손실 함수: binary_crossentropy (이진 분류 문제에 적합)
- 최적화 알고리즘: Adam (적응적 학습률과 모멘텀을 결합해 빠르고 안정적인 학습 지원)
- 학습률: 1e-4 (0.0001)
- 배치 크기: 32
- 에포크 수: 최대 30회
- 데이터 증강: 이미지에 랜덤 뒤집기, 회전, 줌 적용해 과적합 방지 및 일반화 능력 향상
- 환경: TensorFlow 2.x 및 Keras 기반, CPU 계산 환경

04

학습 결과 및 성능 평가 (Results & Evaluation)

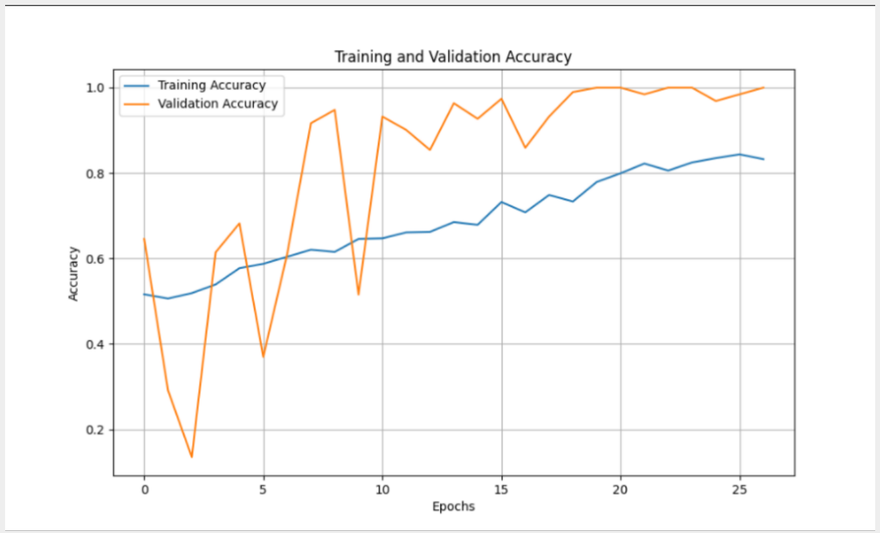
1. 학습 과정 시각화
2. 최종 성능 지표
3. 상세 분석



학습 과정 시각화 - 손실(Loss) 그래프

- 훈련 손실: 점진적 감소 후 0.4 근처 안정화
- 검증 손실: 0.25까지 점진적 감소
- 손실 추세: 두 손실 모두 감소
- 학습 진행: 모델 적응 후 오류 확률 감소

학습 과정 시각화 - 정확(Accuracy) 그래프



- 훈련 정확도: 지속 상승 → 정확도: 약 83%
- 검증 정확도: 변동 있음 → 최종 100% 근접
- 정확도 추세: 검증 정확도가 훈련 정확도를 상회,
- 일반화 우수
- 학습 성능: 모델이 검증 데이터도 효과적으로 분류

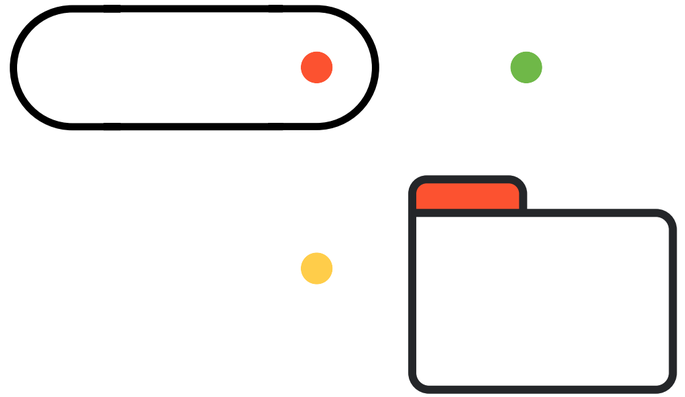
05

결론 및 향후 계획 (Conclusion & Future Work)

1. 프로젝트 요약 및 의의

2. 한계점

3. 향후 계



프로젝트 요약 및 의의

프로젝트 요약 및 의의

- 멀티모달 AI 구현
  (이미지 + 텍스트)
- 멀티모달 학습 가능성 확인
- 이미지 검색, 자동 태깅 등
  응용 기반 마련

한계점

- 데이터 양 제한
- 일반화 성능 개선 필요
- 라벨 불균형
- 데이터 다양성 부족

향후 계획

- 데이터셋 확대 및 다양화
- 융합 방법 연구 강화
- 프로젝트 적용 및
  피드백 반영