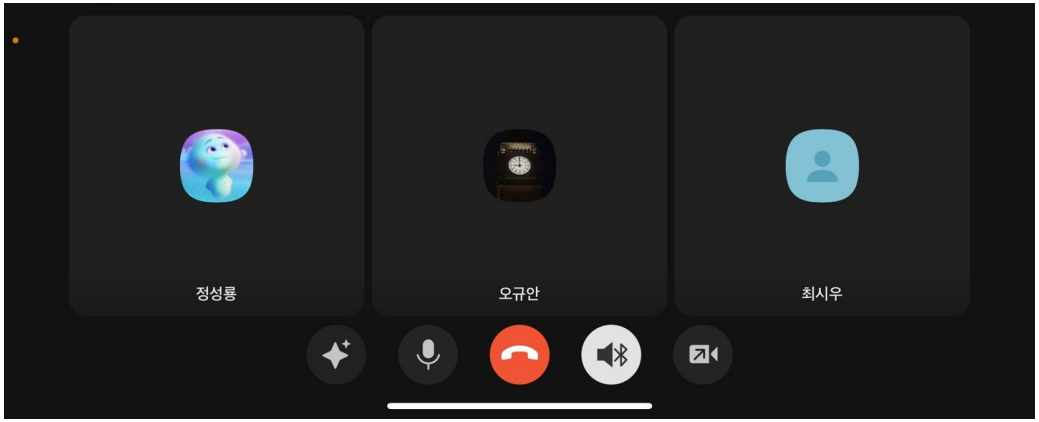
CUAI Advanced Multimodal 프로젝트 02팀 중간 발표

2024.10.01

발표자 : 오규안

프로젝트 팀원 소개



스터디원 1 : 오규안 (AI학과)

스터디원 2 : 최시우 (AI학과)

스터디원 3 : 정성룡 (AI학과)

주제 선정

“VQA (Visual Question Answering)”

Who is wearing glasses?

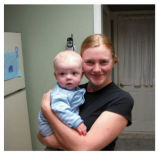
man

woman

Where is the child sitting?

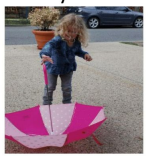
fridge

arms

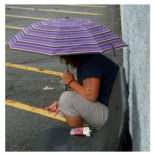


Is the umbrella upside down?

yes



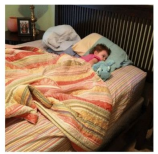
no



How many children are in the bed?

2

1



What is the mustache made of?

AI System

bananas

주제 선정

“Image와 Text 간의 상관관계를 학습하여, VQA의 맥락 이해 성능 향상”

연구 목표:

- **이미지의 특징 영역과 텍스트의 특정 부분 간의 상관관계를** 더 잘 학습할 수 있는 멀티모달 모델을 개발
- Text와 Image의 깊이 있는 상관 관계를 학습하여, **복잡한 질문 맥락** 을 더 잘 이해하는 VQA 시스템 구현
- 이러한 접근을 통해 **VQA 시스템의 성능** 을 단순한 질문뿐만 아니라, 추론이 필요한 복잡한 질문에서도 향상시키는 것을 목표로 함.

세부 연구 내용

1. 멀티모달 트랜스포머 개선

- 기존 트랜스포머 구조를 활용하여, 이미지와 텍스트 간의 상관관계를 더 잘 모델링할 수 있는 멀티모달 트랜스포머 개선
- 이미지에서 중요한 부분을 인식하고, 텍스트에서의 관련된 부분과 연결하는 Cross-Attention Mechanism 강화

세부 연구 내용

2. 이미지와 텍스트의 세밀한 상관 관계 학습

- 이미지의 객체와 텍스트의 명사 또는 형용사 사이의 연관성을 학습하는 방법 개발
- 예) “파란색 공을 들고 있는 사람은 누구인가?” 같은 질문에서, 이미지에서 “파란 공”을 찾아내고, 질문과 연관된 맥락을 더 깊이 이해하는 시스템

세부 연구 내용

3. 부분 객체에 대한 질문 이해 강화

- 질문이 이미지의 일부 객체와 관련된 경우에도 시스템이 정확하게 답할 수 있도록 세부 객체 인식 성능을 강화
- 예) 이미지에서 특정 객체나 사람을 더 정확하게 식별하고, 해당 객체와 관련된 질문에 대한 답변을 개선하는 방식

세부 연구 내용

4. 모델 융합 및 성능 비교

- 기존의 멀티모달 모델 (LXMERT, UNITER 등)과 제안하는 개선된 모델 간의 성능 비교
- 다양한 질문 유형 (사실적 질문, 추론적 질문, 시각적 관계 질문 등)에 대한 성능 차이를 실험적으로 검증

세부 연구 내용

5. 데이터 부족 시 대응

- 모델이 소량의 데이터로도 일반화할 수 있는지 평가하고, Few-shot 학습 기법을 적용하여, 성능을 높이는 연구
- 다양한 상황에서 모델이 적은 데이터로도 적절한 답변을 제공할 수 있도록 하기 위한 전략 검토

기대 결과

- 이미지와 텍스트 간의 상관관계를 더 잘 학습한 모델은 기존의 VQA 모델보다 복잡한 질문에서도 더 높은 정확도를 보여줄 것으로 예상.
- 특히, 이미지 내에서 특정 객체와 관련된 질문에 대한 정확한 답변을 제공하는 성능이 크게 향상될 것으로 기대
- 이 연구는 향후 VQA 시스템의 응용 범위를 확장하고, 보다 자연스럽고 직관적인 인간-컴퓨터 상호작용을 가능하게 할 수 있음.

감사합니다