

AI Eyes

Dec. 2021

Contents

- 2021년 연구개발 톺아보기
- 2022년 연구개발 계획

2021년 연구개발 토크아보기

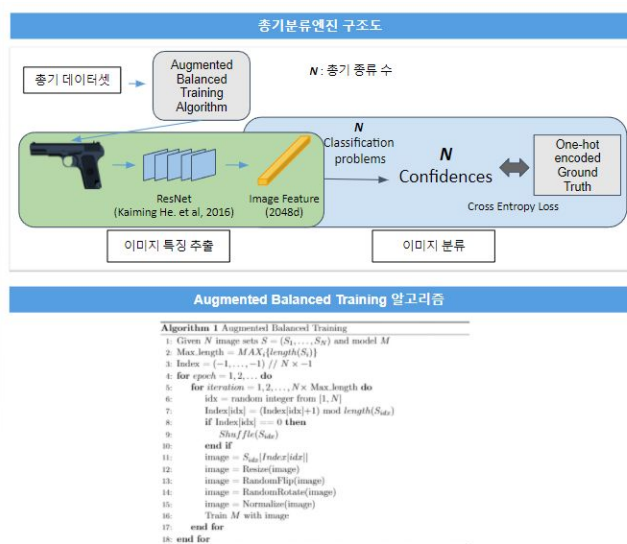
2021년 연구개발 톺아보기

- 과거
- 목표
- 결과

과거

과거

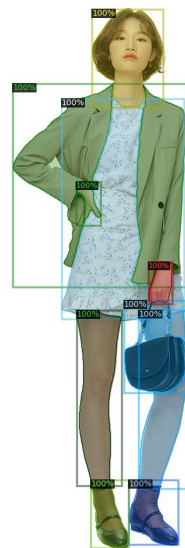
기초적인 시각인식기술 보유



TS팀과 진행한 총기분류엔진



데이터팀과 함께한
이사 견적을 위한 이사짐 검출
(에어뉴)



WUSINSA

의류 segmentation
(스타일봇)

목표

목표

인식을 넘어서 이해하는 기술이 변별력이 될 것

목표

인식을 넘어서 **이해**하는 기술이 변별력이 될 것

장면을 이해하는 기술확보하자!

결과

결 과

장면이해 연구개발

기초시각인식 연구개발

결과

장면이해 연구개발

- 장면에 대한 질의응답이 가능한 Visual Question Answering (VQA)
- 사람과 물체의 상호작용 이해 Human-Object Interaction (HOI)
- 장면을 설명하는 Image Captioning
- 이상행동감지/신원 추적 Person Re-Identification

기초시각인식 연구개발

- 엔진 유지/보수/업그레이드
- 데이터 확보/정제

장면이해 연구개발

- 장면에 대한 질의응답이 가능한 Visual Question Answering (VQA)
- 사람과 물체의 상호작용 이해 Human-Object Interaction (HOI)
- 장면을 설명하는 Image Captioning
- 이상행동감지/신원 추적 Person Re-Identification

장면이해 연구개발

- 장면에 대한 질의응답이 가능한 Visual Question Answering (VQA)
- 사람과 물체의 상호작용 이해 Human-Object Interaction (HOI)
- 장면을 설명하는 Image Captioning
- 이상행동감지/신원 추적 Person Re-Identification

VQA Engine

VQA Engine

- 엔진 소개
- Generative VQA
- OCR based VQA
- Video Question Answering

VQA Engine

- 엔진 소개
- Generative VQA
- OCR based VQA
- Video Question Answering

VQA Engine: 엔진 소개



질문 (번역된 질문) → 대답

사진에 사람이 있습니까? (Do you have anyone in the picture?)

→ yes

사람은 몇 살 처럼 보여? (How old do you look like a person?)

→ 30

사람이 무언가를 들고 있어? (Are you holding something?)

→ yes

사람이 무엇을 들고 있어? (What is a person holding it?)

→ mouse

사람은 뭘 하고 있어? (What are people doing?)

→ working

VQA Engine: 엔진 소개



질문 (번역된 질문) → 대답

사진에 사람이 있습니까?(Do you have anyone in the picture?)

→ yes

사람은 몇 살 처럼 보여?(How old do you look like a person?)

→ 30

사람이 무언가를 들고 있어?(Are you holding something?)

→ no

사람이 무엇을 들고 있어?(What is a person holding it?)

→ laptop

사람은 뭘 하고 있어?(What are people doing?)

→ working

VQA Engine

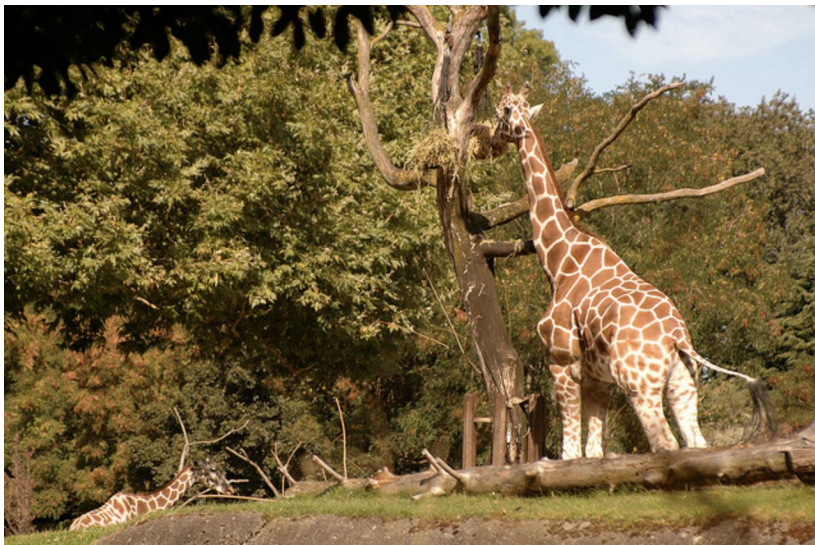
- 엔진 소개
- **Generative VQA**
- OCR based VQA
- Video Question Answering

Generative VQA Engine

학습한 답변만 구사할수 있는 기존의 엔진에서 더 나아가
스스로 대답을 생성할 수 있는 **generative** 엔진

Generative VQA Engine

- 질문: “How many giraffes are there?”

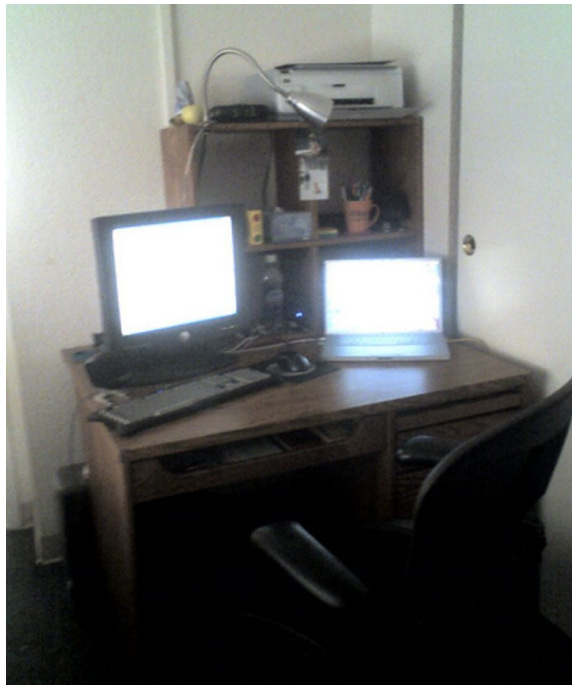


```
[134]: 1 output = model.generate(  
2         input_ids=test_input_id,  
3         vis_inputs=(test_features, test_boxes)  
4     )  
5  
6     print(tokenizer.batch_decode(output))
```

```
['<pad> 2. Two giraffes are eating leaves off a tree.</s>']
```

Generative VQA Engine

- 질문: “What is on the table?”



```
1 output = model.generate(  
2         input_ids=test_input_id,  
3         vis_inputs=(test_features, test_boxes)  
4     )  
5  
6 print(tokenizer.batch_decode(output))
```

```
[ '<pad> laptop. A laptop computer sitting on top of a desk.</s>' ]
```

VQA Engine

- 엔진 소개
- Generative VQA
- OCR based VQA
- Video Question Answering

OCR based VQA Engine

Q: what is the brand of the typewriter?

A: olympia

- OCR detection의 결과인 olympia를 대답
 - 많은 글자 중 olympia가 brand임을 인식



OCR based VQA Engine

Q: what is the name of this library?

A: marshalltown public library

- 건물 (library)의 맥락 이해
- 여러 줄로 이뤄진 글자 감지



VQA Engine

- 엔진 소개
- Generative VQA
- OCR based VQA
- Video Question Answering

Video Question Answering Engine



Caitlin, are you gonna be a little helper on this challenge?

What is the name of the younger girl?

A. Caitlin

C. Jane

B. Lucy

D. Cindy

How many people are playing?

A. 2

C. 3

B. 4

D. 1

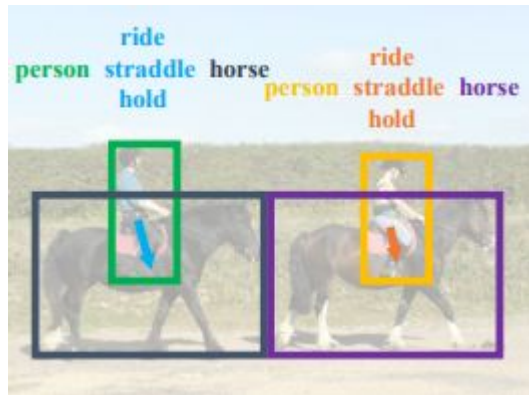
장면이해 연구개발

- 장면에 대한 질의응답이 가능한 Visual Question Answering (VQA)
- 사람과 물체의 상호작용 이해 Human-Object Interaction (HOI)
- 장면을 설명하는 Image Captioning
- 이상행동감지/신원 추적 Person Re-Identification

Human-Object Interaction

“ATM기 앞에서 수상한 사람 검출”과 같은 문제를 풀 때,

VQA보다 더 간단한 데이터셋으로 해결가능



장면이해 연구개발

- 장면에 대한 질의응답이 가능한 Visual Question Answering (VQA)
- 사람과 물체의 상호작용 이해 Human-Object Interaction (HOI)
- 장면을 설명하는 Image Captioning
- 이상행동감지/신원 추적 Person Re-Identification

Image Captioning



The man at bat readies to swing at the pitch while the umpire looks on.



A large bus sitting next to a very tall building.



A horse carrying a large load of hay and two people sitting on it.



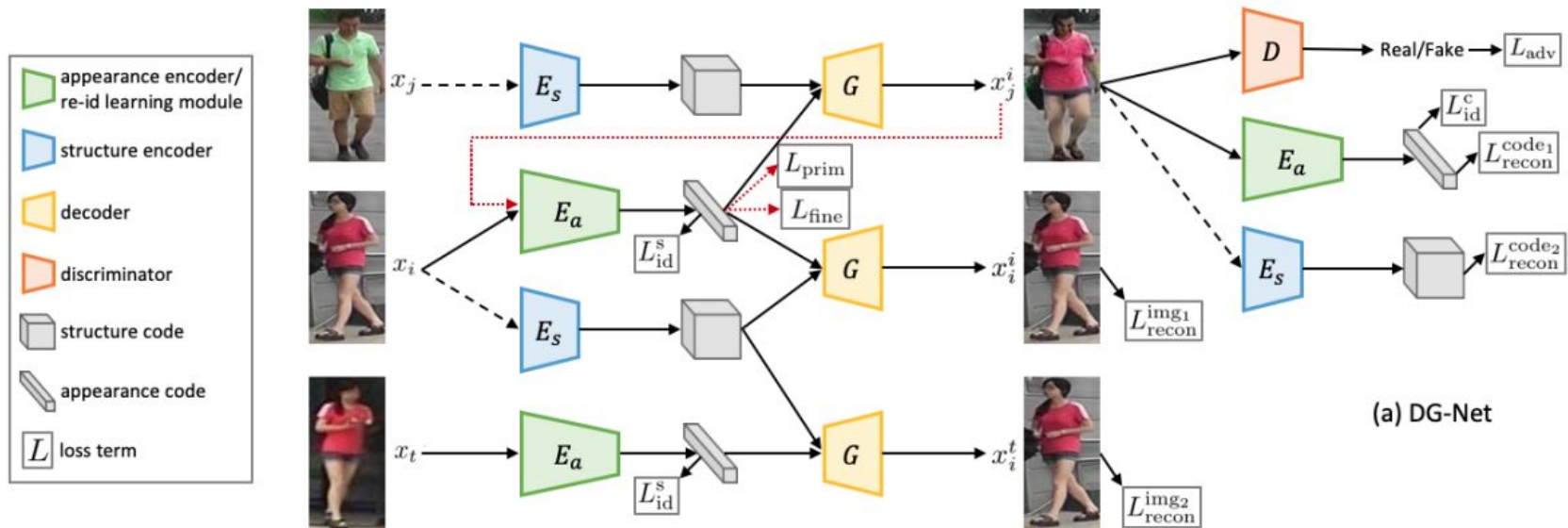
Bunk bed with a narrow shelf sitting underneath it.

장면이해 연구개발

- 장면에 대한 질의응답이 가능한 Visual Question Answering (VQA)
- 사람과 물체의 상호작용 이해 Human-Object Interaction (HOI)
- 장면을 설명하는 Image Captioning
- 이상행동감지/신원 추적 **Person Re-Identification**

Person Re-Identification

알고리즘: DG-Net (CVPR'19)



(a) DG-Net

결과

기초시각인식 연구개발

장면이해 연구개발 **<= 차별점!!**

2022년 연구개발 계획

2022년 연구개발 계획

기초시각인식 연구개발

동영상인식 연구개발 **New!!**

장면이해 연구개발

동영상이해 연구개발 **New!!**