

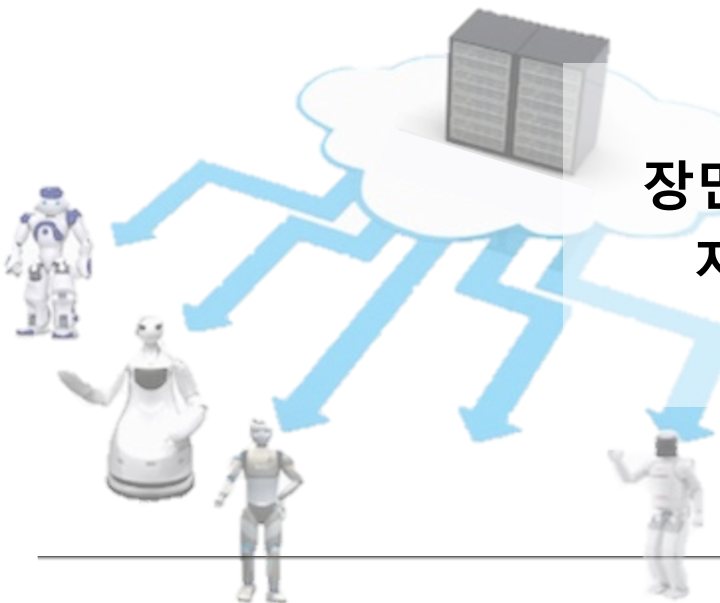
클라우드 로봇 지능 + 열린 세계와 지속 성장 패러다임

2022.6.28

장민수 (mins@etri.re.kr)

지능로보틱스연구본부

한국전자통신연구원



목 차

1. 클라우드 로봇 지능
2. 열린 세계 인지와 지속 성장
3. 정리



클라우드 로봇 지능

서비스 로봇의 현실

IT·통합 ▾

현장 투입된 감정인식 로봇들 해고 위기 "재계약안해"

폐쇄 로봇 해고 위기, 기업용 서비스 계약 연장 의향 15% 불과

주간기술동향 최신ICT이슈/로봇 | 2018. 10. 24. 10:34 | Posted by 보라개구리 Purple Frog

뉴스 2019년 03월 06일 17시 59분 KST | 업데이트됨 2019년 03월 07일 10시 12분 KST

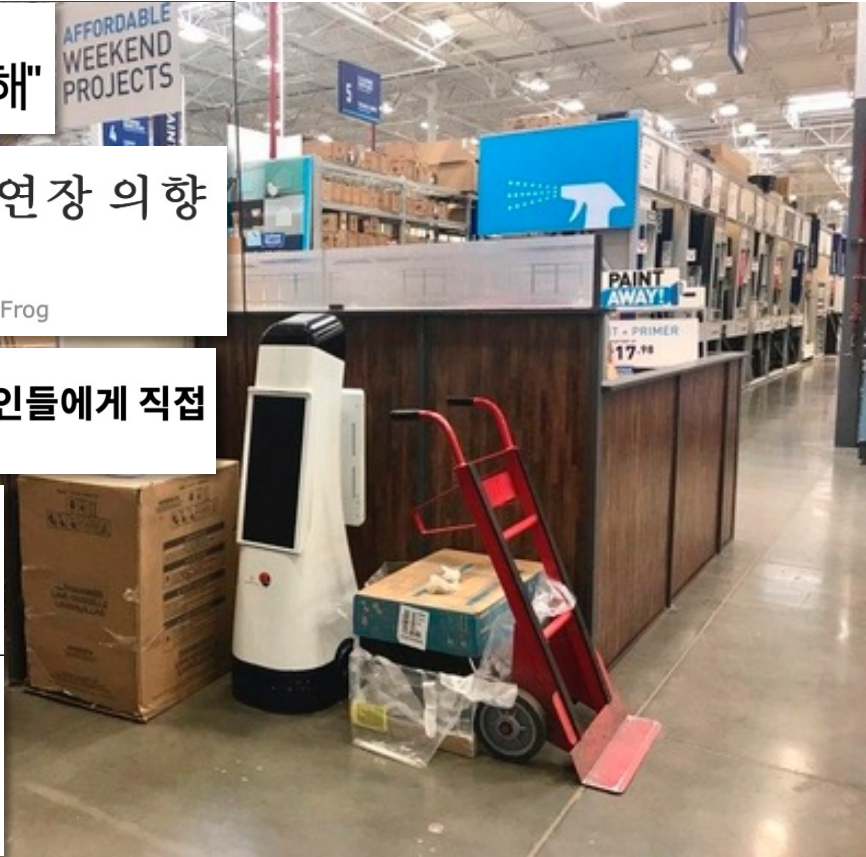
'사회적인 로봇의 외로운 죽음' - 최초의 소셜 로봇 지보가 주인들에게 직접 자신의 죽음을 전했다(영상)

폐기되는 '인공 바보들(Artificial Dumbs)'

조선비즈 | 새너제이, 샌프란시스코=황민규 기자

입력 2017.08.14 06:00

로봇 개발 과정에서 인간과의 소통에 방점이 찍히는 이유는 최근 수년간 세계 각지에서 폐기되는 로봇의 수가 늘고 있다는 점과 무관하지 않다. 각종 로봇이 미디어의 스포트라이트를 받으며 등장했지만, 인간과의 소통 문제로 '스위치 오프(전원 장치 끄)' 운명에 처하고 있는 것이다.



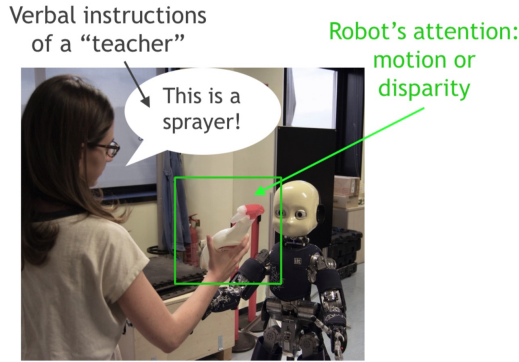
서비스 로봇의 한계

- **현재 서비스 로봇 기술은 변화하는 환경에서도 같은 수준의 상호작용을 제공**
 - 상업적 용도의 서비스 로봇에서 활용이 제한
 - **환경과 사용자 변화에 적응하는 맞춤형 서비스를 제공함으로써** 다양한 서비스 분야에서 신시장 창출 가능
 - ※ 데이터·AI 경제 활성화 계획('19.1.), AI R&D 전략('18.5.) 및 혁신성장 실현을 위한 5G+ 전략('19.4) 등에서 **AI·로봇·클라우드가 융합된 新산업 육성 계획 발표**
- **상업용·개인용 서비스 로봇은 높은 수준의 상호작용, 맥락이해, 개인화 기술 필요**
 - 현재 서비스 로봇의 멀티모달 상호작용과 맥락 이해 기술은 다수 사용자 환경에 공통 적용을 전제로 한 범용 지능에 의존
 - **맥락 기반 개인화 및 적응화 기술**은 다양한 상황에서 효과적으로 로봇 서비스를 제공할 수 있음
- **맥락이해와 지역적 적응 학습 기술로 기존 서비스 로봇의 상업적 경쟁력 제고 가능**
 - **클라우드 로봇 지능을 통해 자체 기술력이 부족한 중소기업과 스타트업의 로봇 서비스 수준 향상**

Domain Shift in Robot Vision

Object Detection with robot vision (Angeletti et al., 2018*)

Human-Robot Interaction Context



Domain Shift by Translation and Scale



Performance downgrades by domain shift

	S	T	$S \rightarrow S$	$S \rightarrow T$
translation	left	right	98.33	45.80
	right	left	99.33	54.49






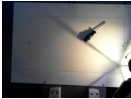
	S	T	$S \rightarrow S$	$S \rightarrow T$
scale	close	far	99.45	18.44
	far	close	98.67	28.80

*Angeletti, Gabriele, Barbara Caputo, and Tatiana Tommasi. "Adaptive deep learning through visual domain localization." In 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 7135-7142. IEEE, 2018.

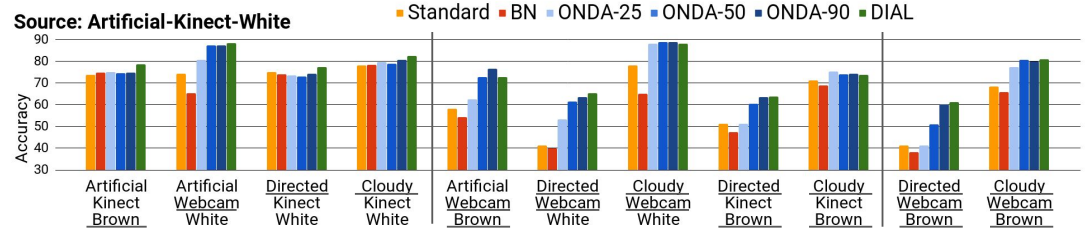
Domain Shift in Robot Vision

Object Recognition with robot vision (Massimiliano et al., 2018*)

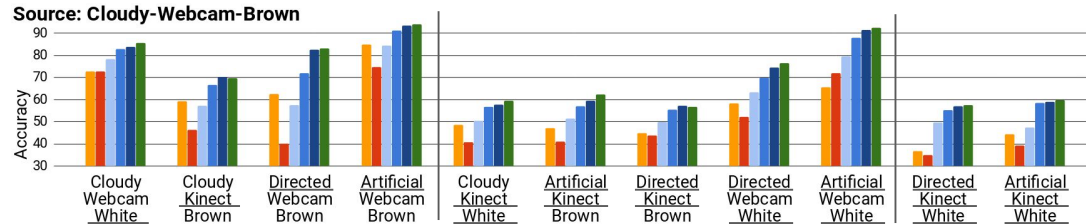
Domains

Camera Type	Illumination		
	Artificial	Cloudy	Directed
Kinect			
Webcam			

Accuracy on different domains



(a) Source Domain: Artificial light, Kinect camera and White background

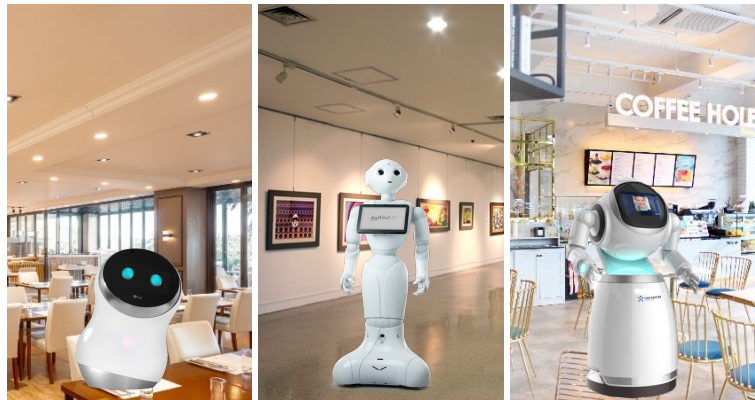
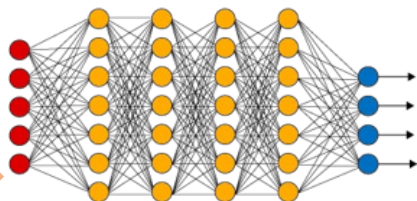


*Mancini, Massimiliano, Hakan Karaoguz, Elisa Ricci, Patric Jensfelt, and Barbara Caputo. "Kitting in the wild through online domain adaptation." In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1103-1109. IEEE, 2018.

로봇의 맥락 이해 지능

현재 서비스 로봇의 맥락 이해 지능은.....

정적, 폐쇄적



대용량 데이터셋 구축
Labeled Dataset

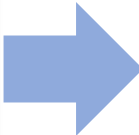
훈련

배포 활용

로봇의 맥락 이해 지능

현재 서비스 로봇의 맥락 이해 지능은.....

정적, 폐쇄적



서비스 환경 변화

태스크 변화

검출/인식 성능 저하

검출/인식 불가능

클라우드 로봇 지능



도메인 적응: Unsupervised & Active

연구 목표

- 지역 도메인의 특성을 활용하여 모델의 성능을 향상시키는 기술(Domain Adaptation)을 활용한 서비스 맥락 이해 기술의 고도화
- 사용자 참여를 통해 지역 도메인의 지식을 증강시키고 이를 활용하는 Interactive Target Domain 적응 기술 개발

핵심 연구개발 결과

- 학습 및 실제 환경 사이 차이로 인한 성능 저하 확인
- 비지도 도메인 적응 (UDA) 기술 개발 및 성능 개선 확인
- 사용자 참여를 통한 지식 증강 기술 적용 및 성능 개선 확인

• 비지도 도메인 적응 기술 개발

- 시점 변화(90도 → 10도) 시 83.39% → 67.48% 성능 저하
- 스타일 적응⁽¹⁾, 모조 정답 부여, 특징 정렬 기법⁽²⁾ 적용
- 음식 검출, 양 인식, 진행 상태 4%p ~ 9%p 성능 개선

• 능동적 도메인 적응 기술 개발

- 지역 데이터 중 다양성과 불확실성 기반 샘플 선택
- 987개 지역 샘플 중 100/300/500개를 샘플로 선택
- 소수 샘플의 경우 적응 성능: 불확실성 > 다양성 > 무작위
- 100장 샘플 활용 시 정확도 12%p ~ 13%p 개선

자체 구축 데이터 예



촬영각 도메인 차이:
왼쪽부터 90도, 45도, 10도

도메인 차이에 따른 성능 저하

촬영각도	3종 평균
All → All	81.91
10 → 10	83.39
45 → 10	85.65
90 → 10	67.48

자체 DB에서 도메인 적응기술 적용 전후 인식 성능 비교

UDA기반	source-only	특징 정렬	스타일 적응	모조 정답	스타일 적응+모조 정답
평균	67.48	67.12	67.76	70.25	73.30 (+5.82)

ADA기반	source-only	무작위	다양성	불확실성	target-only
평균	67.48	77.09/80.21/ 82.51	77.40/81.47/ 82.50	80.02/81.66/ 82.51	83.39

도메인 적응: Unsupervised & Active

도메인 적응에 따른 서비스 맥락 이해 지능 성능 비교

도메인 적응 전
초기 모델을 이용한 인식 결과



음식 검출과 양 추정 오류가 다수 발생

도메인 적응 후,
인식 결과



음식 검출과 양 추정이 상대적으로 정확

도메인 적응: Active

연구 목표

- 초기 학습 환경과 상이한 지역 도메인으로의 적응을 위한 신경망 구조와 Unsupervised Domain Adaptation 기법
- 도메인 별로 특화된 특성과 도메인 사이에서 공유하는 특성을 분리하여 효과적으로 학습하기 위한 신경망 계층(Layer)

핵심 연구개발 결과

- 정답 확보로 인한 정보 확장 효율을 높이기 위한 능동 학습 도메인 적응 방법을 제안
- DomainNet 데이터셋 대상 상이한 도메인 대상 분류 정확도 59.1% 달성 (3-shot Active Domain Adaptation, SOTA 성능)

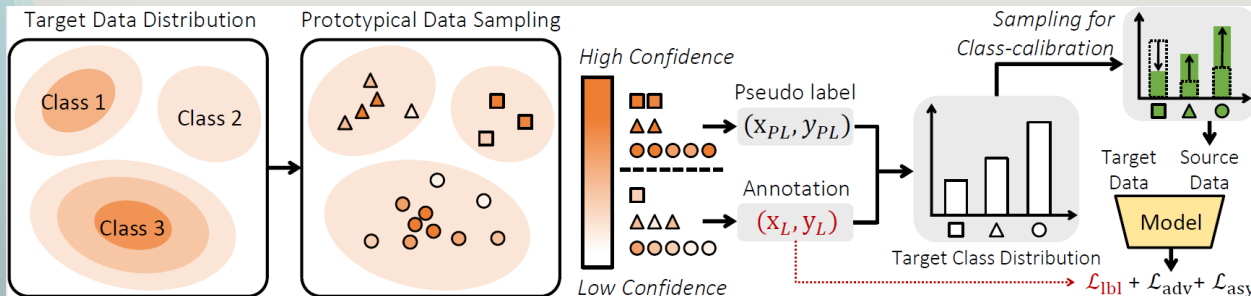


지역 획득 데이터 중 로봇이 선별한 소수 샘플의 정답을 확보하여 활용

DB: 사물 345종, 도메인 6개, 60만장

성능

기술	정확도(%) (1-shot)	정확도(%) (3-shot)
MME	44.2	48.2
APE	44.6	48.9
MetaMME	48.8	-
BiAT	45.5	49.4
CDAC	52.1	56.2
Ours	55.1	59.1



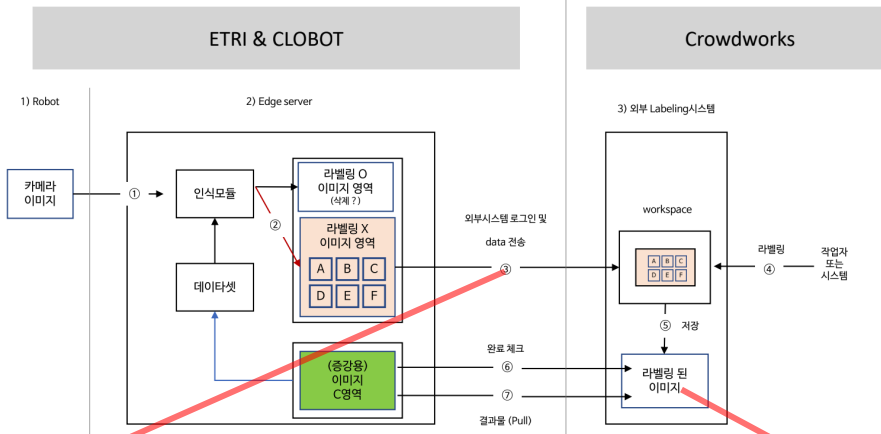
기술 열개 (CVPR'22 논문 투고)

도메인 적응: 온라인 태깅 시스템

개발 목적

- 정답이 없는 지역 획득 데이터에 대한 정답을 확보할 수 있는 Human-in-the-Loop 시스템
- 능동적 도메인 적응 방법 활용에 최적
- 로봇을 통해 입수한 이미지와 지역로봇 지능 모듈이 추론한 결과를 함께 온라인 태깅 시스템에 제공하고 정답 여부 확인, 오류 보정, 미검출 내역 추가를 요청

온라인 태깅 시스템 구조



개발 결과

- 상용 클라우드소싱 기반 데이터 태깅 시스템과 연동
- 태깅 요청 후 응답까지 약 1~2일 소요
- 이미지 2,000장으로 시스템 운영 시험 진행
- 향후 테스트베드와 실환경 기반 기술 검증에 활용 계획

정답 요청 예시

```
{
  "req_id": 1104891,          # 요청 번호
  "raw_data": "img_20210730_01483.jpg", # 원본 데이터 파일 이름
  "task": 1,                 # 알금 레이블링 태스크
  "result": {
    "clothing_dets":        # 지능 모듈 실행 결과
      # 의상 검출/인식 결과
      [
        {
          "bbox": (.....), "style": 'suit', # 사람 영역 좌표, 수트
          "bbox": (.....), "style": 'uniform', # 사람 영역 좌표, 유니폼
          ...
        }
      ]
  }
}
```



담고 있는 정보



의상 스타일: 수트, 의상 스타일: 유니폼

정답 데이터 예시

```
{
  "req_id": 1104891,          # 요청 번호
  "raw_data": "img_20210730_01483.jpg", # 원본 데이터 파일 이름
  "task": 1,                 # 알금 레이블링 태스크
  "result": {
    "clothing_dets":        # 지능 모듈 실행 결과
      # 의상 검출/인식 결과
      [
        {
          "bbox": (.....), "style": 'suit', # 사람 영역 좌표, 수트
          "bbox": (.....), "style": 'uniform', # 사람 영역 좌표, 유니폼
          "bbox": (.....), "style": 'uniform', # 사람 영역 좌표, 유니폼
          "bbox": (.....), "style": 'uniform', # 사람 영역 좌표, 유니폼
          ...
        }
      ]
  }
}
```

빨간색 폰트: 보정된 정답
파란색: 추가된 정답



의상 스타일: 수트, 의상 스타일: 유니폼, 의상 스타일: 유니폼

도메인 적응: 정답 확보 비용 절감

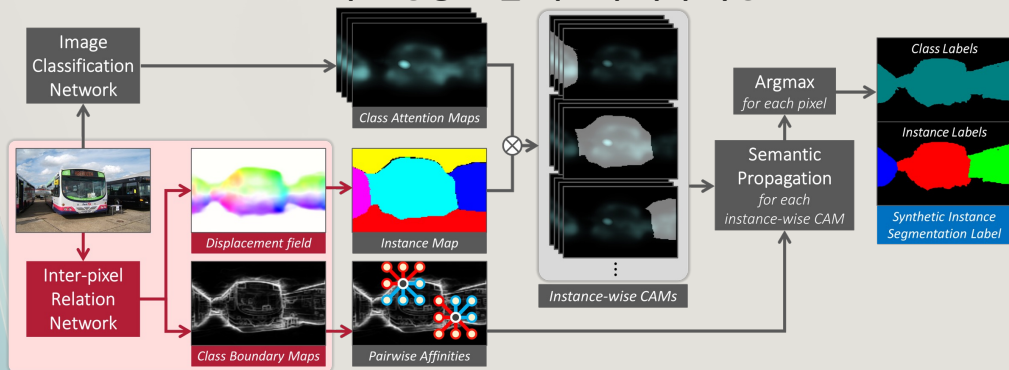
연구 목표

- 사용자 입력을 최소로 받아 지역 획득 데이터의 정답을 생성하는 Weakly Supervised Learning 기술과 신경망
- 소수 정답 데이터를 이용해 지역 획득 데이터의 정답을 생성하는 Few-Shot Learning 기술과 신경망

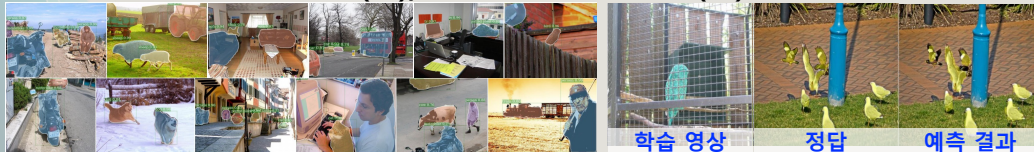
핵심 연구개발 결과

- 이미지 분류 정답에서 의미/개체 분할, 윤곽 검출 정답을 생성하는 WSL 기술 개발: SOTA 성능 달성
- 소수의 의미 분할 정답에서 다수 데이터의 의미 분할 정답을 생성하는 Few-Shot Learning 기술 개발: SOTA 성능 달성
- SCI급 저널/최우수 학회 논문 2편 투고 (IJCV, CVPR 2021)

WSL 기반 영상 분할 기술의 처리 과정



개체 분할(좌), Few-Shot 의미 분할(우) 예시



성능 비교표

(위로부터 의미 분할, 개체 분할, 경계 추출, Few-Shot 의미 분할)

모델	정답 데이터	부가 정보	AP_{50}^r
PRM [3]	이미지 분류	MCG [4]	26.8%
SDI [5]	사물 검출	BSDS [6]	44.8%
SDS [7]	영상 분할	MCG [4]	43.8%
Ours	이미지 분류	-	46.7%

모델	정답 데이터	부가 정보	평균 IoU
DSRG [8]	이미지 분류	MSRA-B [9]	63.2%
SDI [5]	사물 검출	BSDS [6]	67.5%
Ours	이미지 분류	-	64.8%

모델	정답 데이터	부가 정보	평균 IoU
CASENet [13]	이미지 분류	VGG19	56.8%
Det+HED [14]	사물 검출	VGG16	51.9%
Ours	이미지 분류	Resnet50	59.8%

모델	1-shot mIoU	5-shot mIoU
PANet [15]	48.1%	55.7%
PGNet [16]	56.0%	58.5%
Ours	56.7%	62.6%

도메인 적응: 실환경 데이터셋 확보

외형 특징 데이터

- 서비스 분야: 식당/전시관의 개인화 맥락 이해
- 정답: 의상 속성 (스타일, 색상, 무늬, 소매 길이 등)
- 도메인
 - 장소: 5개 (식당 3곳, 전시관 2곳)
 - 위치: 장소 별 5곳 (조명 등 환경 조건 차이 고려)
 - 시간대: 2종 (낮, 밤)
- 데이터 형식: RGB 동영상
- 규모: 직찰 동영상 5,000개



동일 장소의 다른 위치와 시간대 촬영 영상

식사 서비스 맥락 데이터

- 서비스 분야: 식당의 식사 서비스 맥락 이해
- 정답: 음식 종류, 음식 양, 식사 상태(식사 전/중/후)
- 도메인
 - 음식 종류: 5종(한식, 양식, 일식, 중식, 패스트푸드)
 - 매장: 음식 종류 별 5개
 - 위치: 매장 별로 2곳
 - 시야각: 3종 (10도, 45도, 90도)
- 데이터 형식: RGB 동영상
- 규모: 직찰 동영상 50개 / 유튜브 동영상 70개





열린 세계 인지와 지속 성장

현재 로봇은 닫힌 세계 가정 하에 세상을 인지

• 기존 로봇은 “지능의 유연성” 부재

- 지식의 불확실성을 판단하지 못함
- 새로운 지식을 수용하여 성장하지 못함

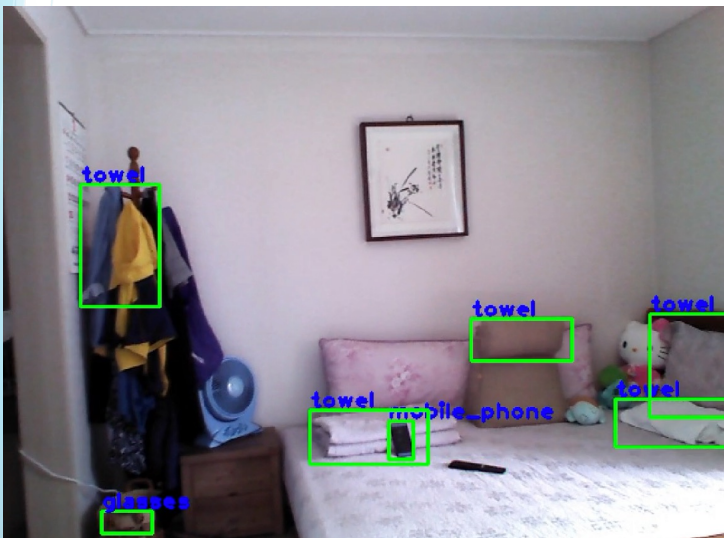


새로운 환경과 데이터에 **적응하지 못함**
 지능이 **성장하지 않음**

AI 활용성 저하

로봇의 인지 문제

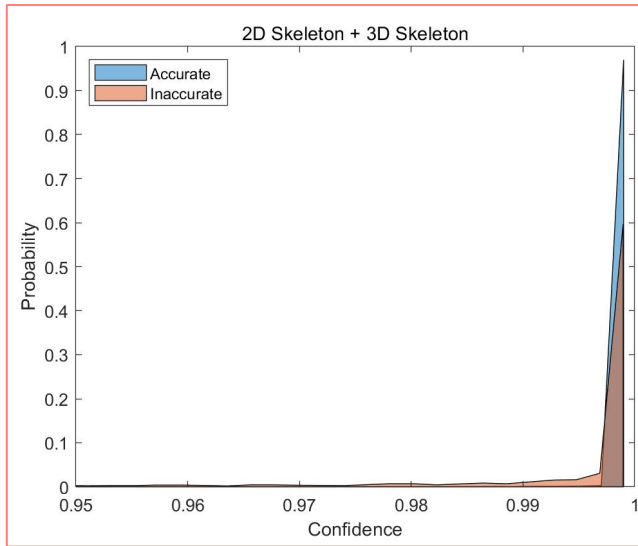
인공지능의 실환경 적용에서 광범위하게 발생



처음 보는 물체들을 오인

로봇 기반 사물 검출 사례 (20종 사물 훈련)

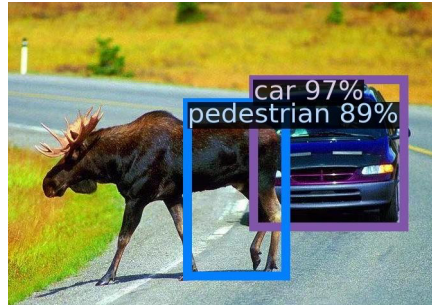
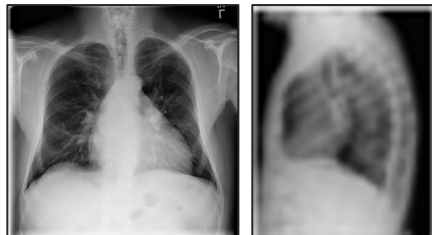
손목 -> 양말, 형광등 -> 핸드폰,
선풍기 -> 모자, 고무장갑 -> 컵 등



Over-Confidence

일상행동인식의 Confidence 분포

대부분의 인식 오류 상황에서 Confidence는 0.99 ~ 1.0 사이에 분포



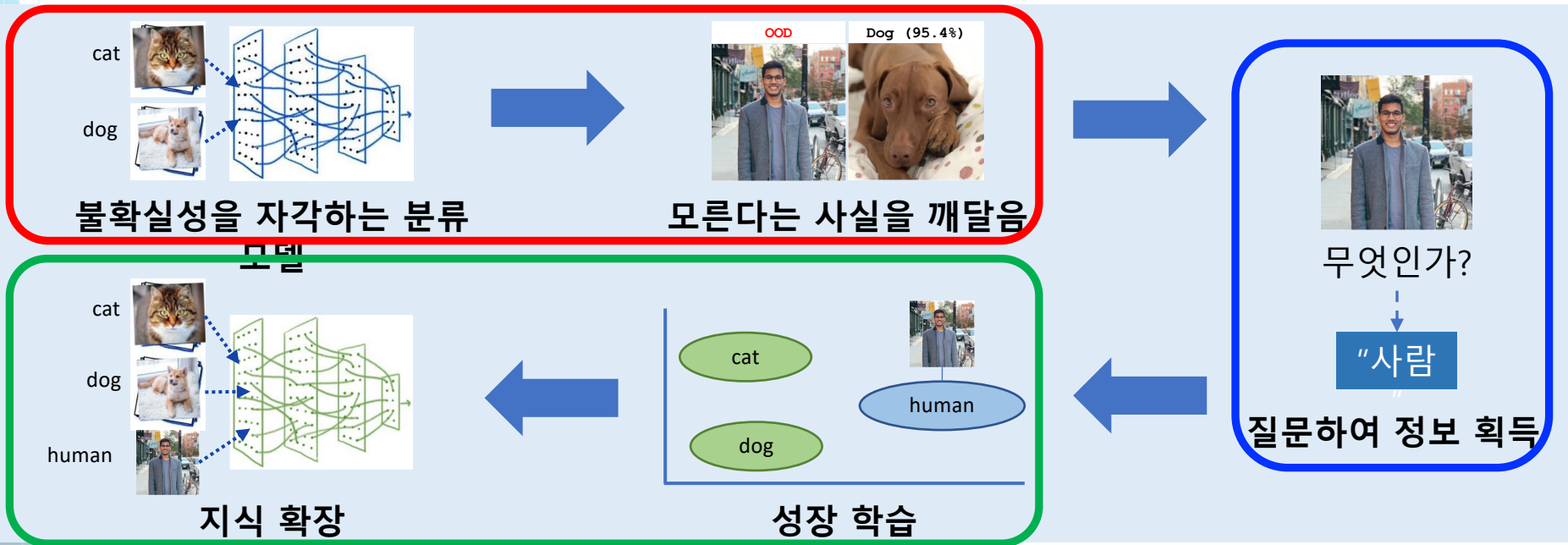
Mission-Critical Cases

의료 진단, 자율차,
안전 관리 등

불확실성 관리 필요

열린 세계 인지와 지속 성장이 필요

- 불확실성 자각과 질의응답을 통해 “지능의 유연성” 확보
 - 학습하지 않은 멀티모달 OOD* 정보를 접하면 불확실성을 자각
 - 질의응답을 통해 불확실성을 해소하고 학습하며 성장 * OOD: Out-Of-Distribution



Out-Of-Distribution Generalization

• 불확실성 자각은 차세대 인공지능의 핵심 원천 기술

- Yoshua Bengio 교수는 에이전트의 학습에 있어 가장 중요한 요소 중 하나로 **OOD 일반화 (OOD Generalization)** 능력을 제기
- 훈련 상황과 상이한 응용 상황에 대응하는데 OOD 일반화는 핵심 요소

AGENT LEARNING NEEDS
OOD GENERALIZATION

Agents face non-stationarities

Changes in distribution due to

- their actions
- actions of other agents
- different places, times, sensors, actuators, goals, policies, etc.



*Multi-agent systems: many changes in distribution
Ood generalization needed for continual learning*

장소, 시간, 센서와 액추에이터, 목표,
정책 등 각종 요인의 변화
+ 에이전트와 타 에이전트의 행위

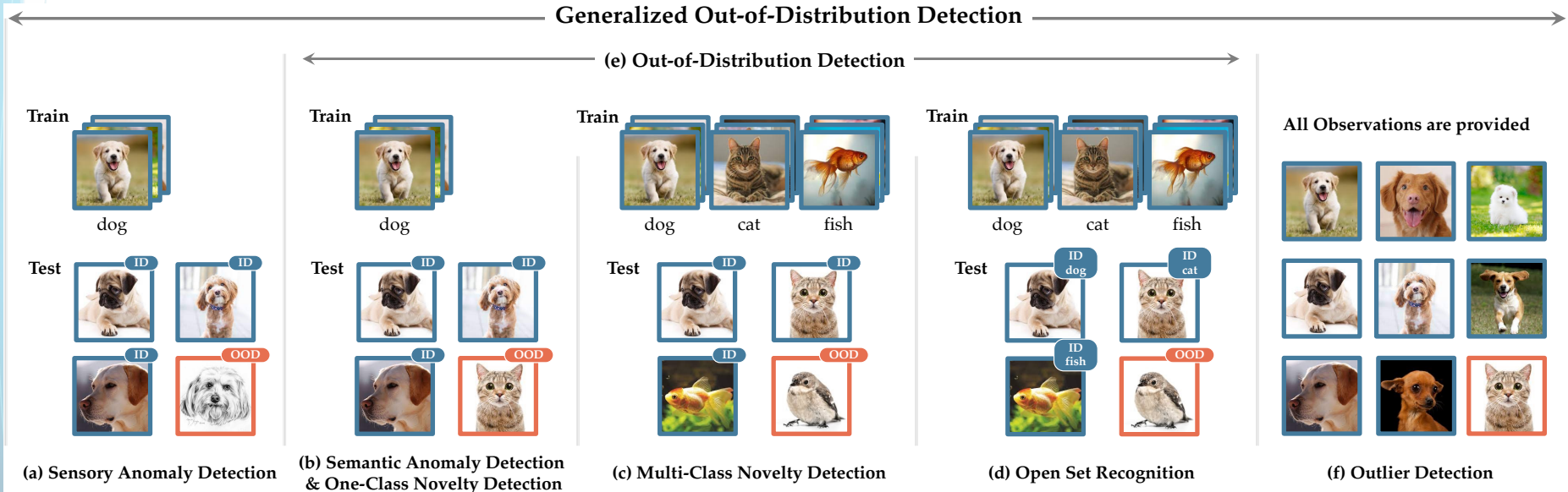
비정상성 (Non-Stationarities)
발생

대응을 위해
OOD 일반화 능력 필요

출처: Yoshua Bengio, "From System 1 Deep Learning to System 2 Deep Learning," NeurIPS Invited Talk, 2019.

모르는 점을 알기

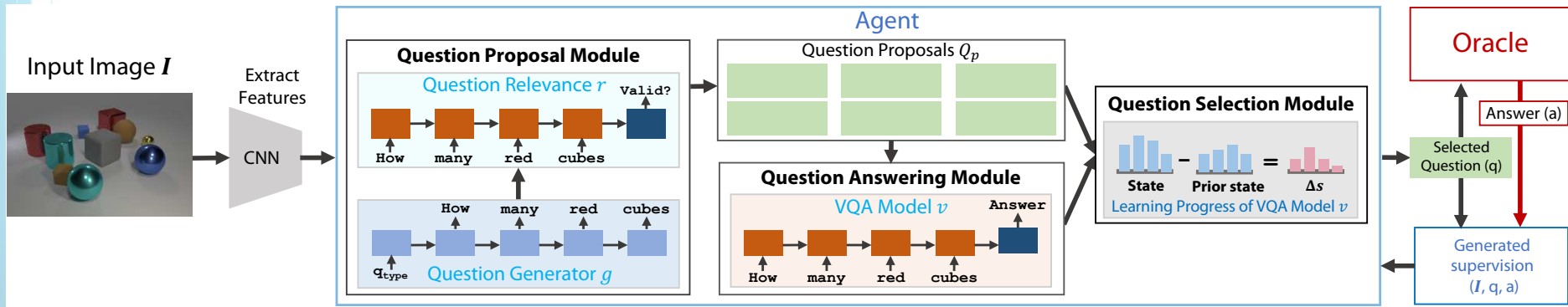
먼저 스스로 무지를 깨우쳐야 성장할 수 있습니다.



Yang, Jingkan, et al. "Generalized out-of-distribution detection: A survey." *arXiv preprint arXiv:2110.11334* (2021).

질문하며 성장하기

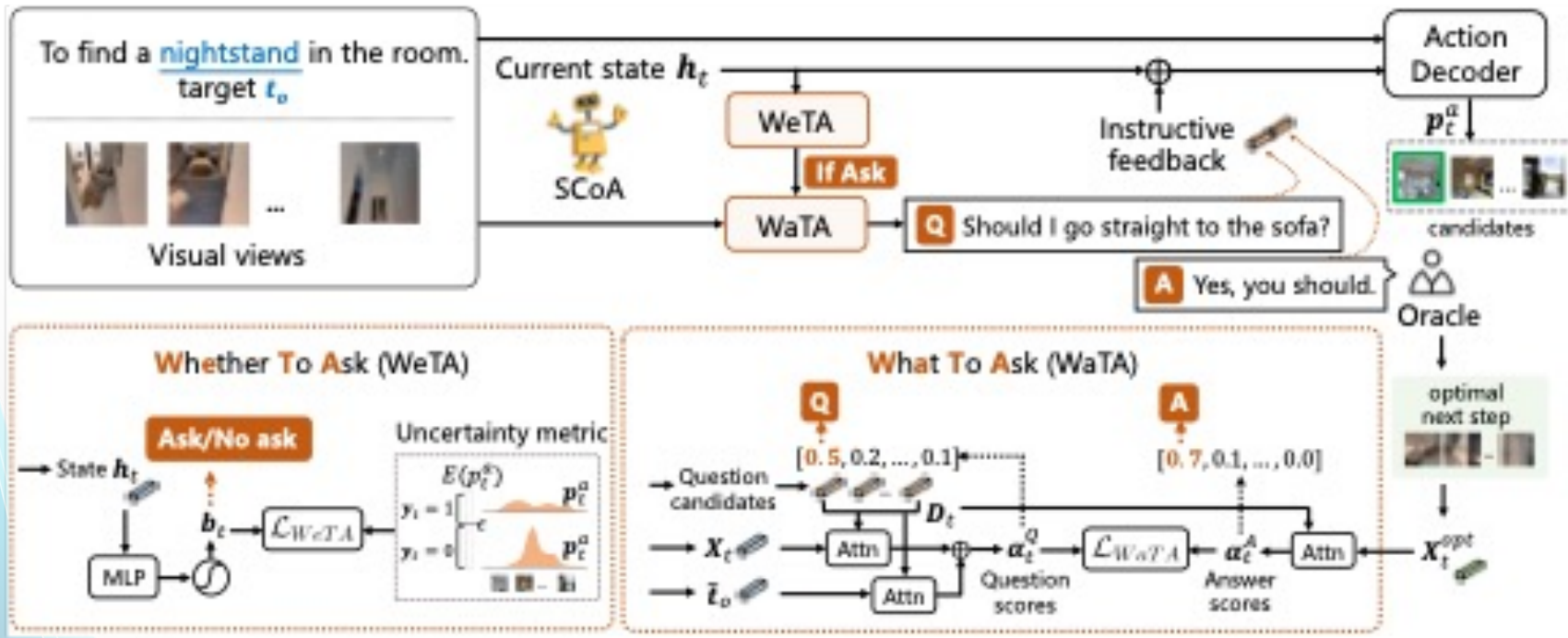
Learning by Asking Questions



Misra, Ishan, et al. "Learning by asking questions." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

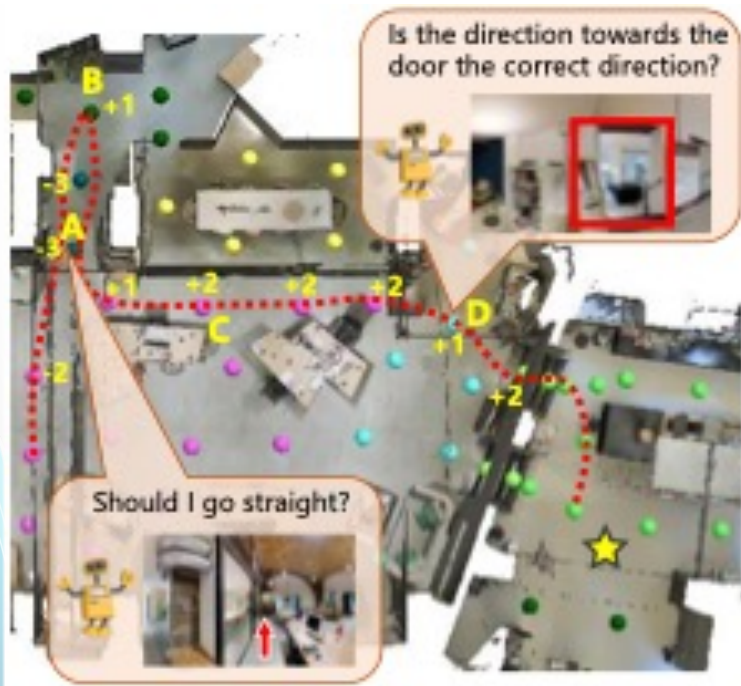
질문하며 성장하기

Visual Dialog Navigation



질문하며 성장하기

Visual Dialog Navigation



Method	Val Unseen		
	SR \uparrow	OSR \uparrow	SPL \uparrow
Random	1.76	11.93	1.01
R2R Teacher Forcing [1]	3.21	4.94	2.80
R2R Student Forcing [1]	12.88	4.20	8.07
RCM [27]	9.29	14.23	6.97
Self-Monitor [15]	8.15	11.28	6.44
FAST-Short [8]	10.08	20.48	6.17
Navigator-Pointer [19]	14.40	28.20	7.19
SCoA (Ours)	16.94	29.29	8.2

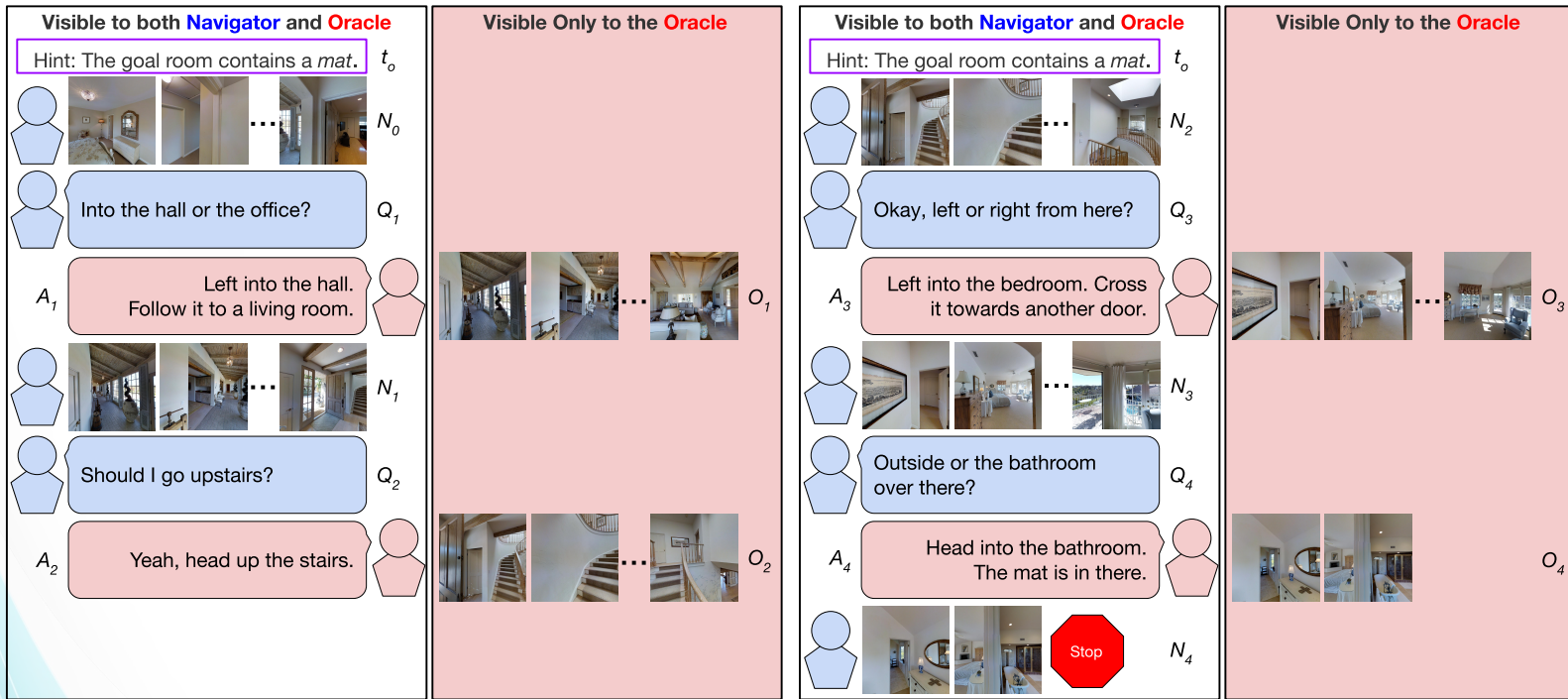
Table 5: Performance comparison on the unseen validation of the REVERIE. Three metrics, including SR (%), OSR (%) and SPL (%), are introduced.

질문하며 성장하기

Vision-and-Dialog Navigation

https://youtu.be/BonlITv_PKw

<https://cvdn.dev>



Thomason, Jesse, Michael Murray, Maya Cakmak, and Luke Zettlemoyer. "Vision-and-dialog navigation." In *Conference on Robot Learning*, pp. 394-406. PMLR, 2020.

Large Language Models...

Robot: Hi there, I'm a robot operating in an office kitchen.

Robot: You can ask me to do various tasks and I'll tell you the sequence of actions I would do to accomplish your task.

Human: How would you hold the snickers?

Robot: 1. pick up the snickers, 2. done.

Human: How would you put the trailmix here?

Robot: 1. put down the trailmix, 2. done.

Human: How would you put a water bottle and an oatmeal next to the microwave?

Robot: 1. find a water bottle, 2. pick up the water bottle, 3. go to microwave, 4. put down the water bottle, 5. find an oatmeal, 6. pick up the oatmeal, 7. go to the microwave, 8. put down the oatmeal, 9. done.

Large Language Models...

Human: How would you put the coke can down on the far counter?

Robot: 1. find a coke can, 2. pick up the coke can, 3. go to the far counter, 4. put down the coke can, 5. done.

Human: I am thirsty. How would you help me?

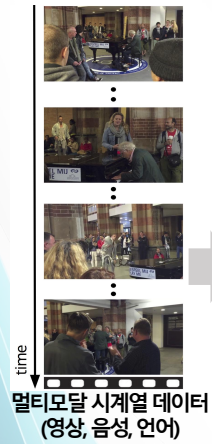
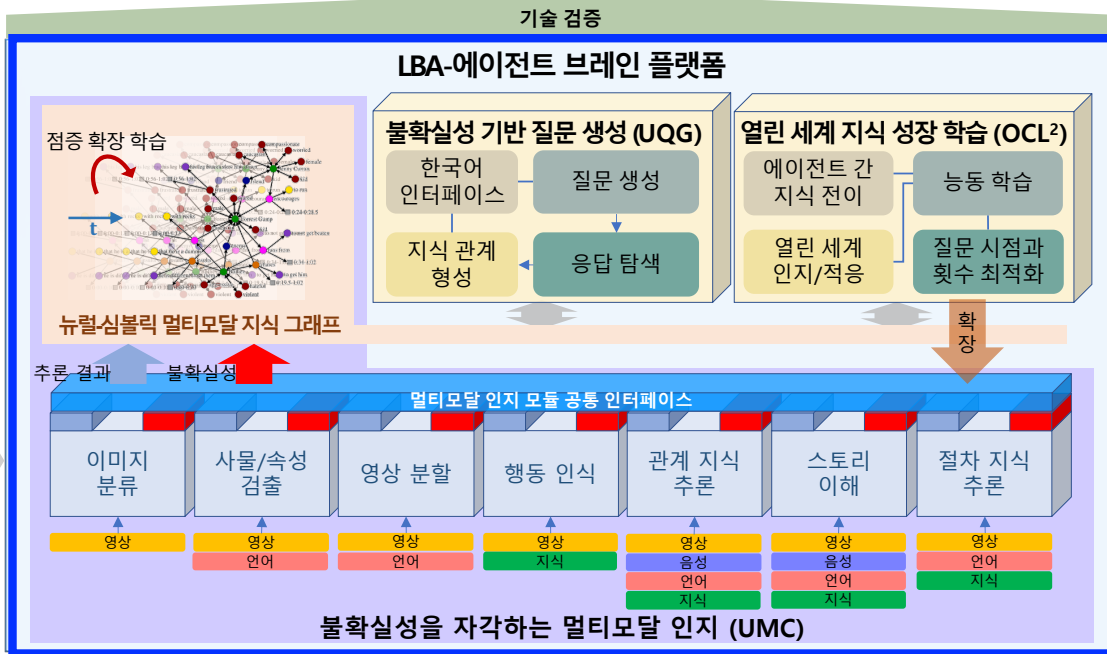
Robot: 1. find a water bottle, 2. pick up the water bottle, 3. bring it to you, 4. put down the water bottle, 5. done.

Large Language Models...

Human: How do you explain the process of making a soup? Robot: I would 1.

1. get a pot, 2. fill it with water, 3. put it on the stove, 4. turn on the stove, 5. wait for the water to boil, 6. add the soup mix, 7. stir, 8. wait for the soup to finish cooking, 9. turn off the stove,

무지를 깨우치고 질문하며 성장하는 AI



정리

정리

- 지금 로봇 지능의 특성은 폐쇄(Closed)되고 고정(Static)됨
- 로봇 운영 환경과 상황은 다양하고, 끊임없이 변화 (Continual Domain Change)
- 클라우드 로봇 지능: 도메인에 지속 적응 (Continual Domain Adaptation)
- 열린 세계 인지와 성장: 무지를 깨닫고 배움 (OOD Detection and Learning-by-Asking)

감사합니다.

실환경에서 잘 동작하는 기술을 넘어,
실환경에서 잘 학습하는 기술을!

An important challenge for machine learning is not necessarily finding solutions that work in the real world, but rather finding stable algorithms that can learn in real world.

(Timothée Lesort, 2020)