

“위성사진 속 건물 하나가 바뀌었다... AI는 알아챌까?”

GIST, 재난·자율주행 등 다양한 상황에서의 AI 문맥 인식 및 영상 이해도 검증 기준 개발

- AI학과 김의환 교수팀, 4개 분야 51개 상황의 4,996개 이미지 쌍으로 구성된 고도화된 벤치마크 개발... 단어 일치도 중심의 기존 방식 탈피해 정확성·구체성·상황 관련성 종합 평가
- GPT-5.2 평가 결과 문장은 자연스럽지만 실제 상황 이해도 및 설명의 일관성은 인간에 크게 뒤쳐져
- 재난 감시·자율주행 등 다양한 AI 성능 개선에 기여 기대... 국제학술대회 'ECCV 2026' 채택



▲ (왼쪽부터) AI학과 김의환 교수, 김재우 석·박사통합과정생

광주과학기술원(GIST·지스트, 총장 임기철)은 AI학과 김의환 교수 연구팀이 두 장의 이미지를 비교해 인공지능(AI)이 주어진 상황에 맞는 중요한 변화를 찾아 정확하게 설명하는지 확인할 수 있는 새로운 평가 기준 ‘C3-Bench(Context-Aware Change Captioning Benchmark, 문맥 인식 변화 설명 평가체계)’를 개발했다고 밝혔다.

AI의 성능을 제대로 비교하려면 같은 문제를 주고 어떤 AI가 더 정확하게 해결하는지 판단할 수 있는 기준이 필요하다. C3-Bench는 두 장의 이미지와 함께 ‘무엇을 살펴볼 것인지’를 제시하고, AI가 그 목적에 맞는 변화를 골라 설명하는지 확인하는 일종의 ‘AI 변화 이해력 시험’이다.

두 사진에서 다른 부분을 많이 찾아낸다고 해서 AI가 사진 속 변화를 잘 이해한다고 할 수는 없다. 무엇을 확인하려는지에 따라 중요하게 봐야 할 변화가 달라지기 때문이다.

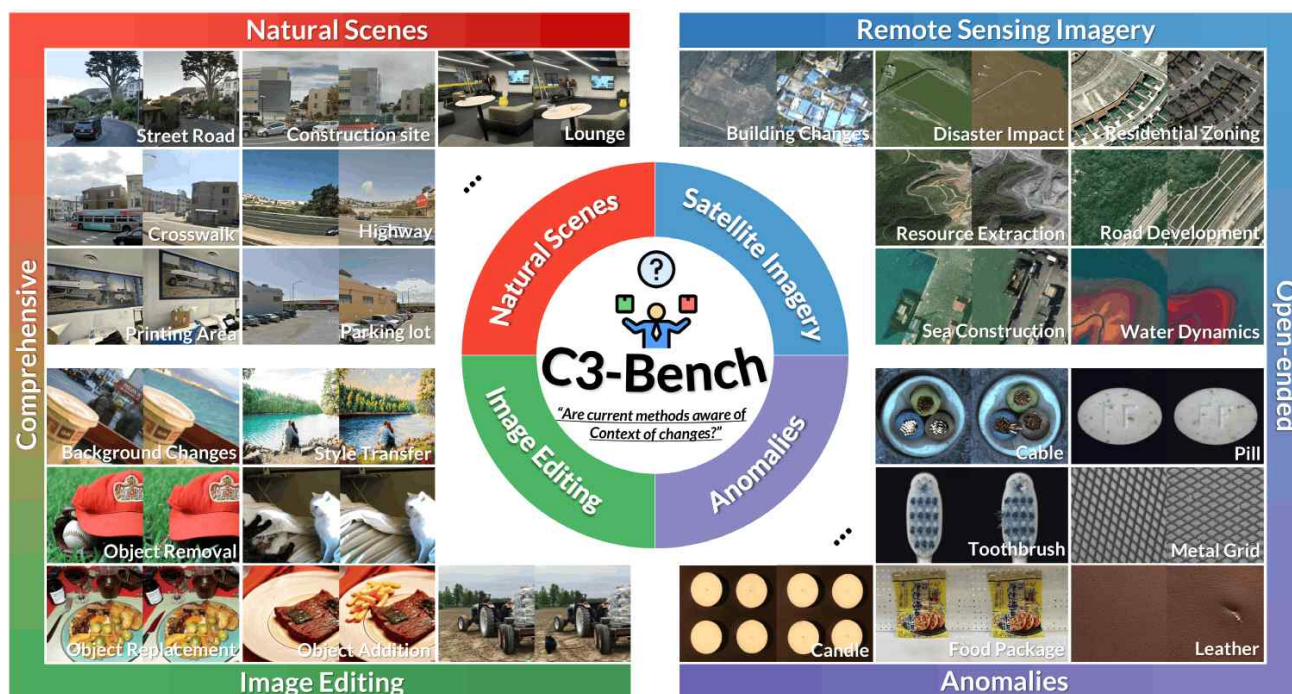
예를 들어 같은 철도 사진 두 장이라도 날씨 변화를 살피는 상황에서는 눈이 쌓이거나 구름이 줄어든 것이 중요한 변화가 될 수 있다. 반면 철도 안전을 확인하는 상황에서는 선로에 새로 나타난 열차가 핵심적인 변화이고, 날씨 차이는 중요하지 않을 수 있다.



▲ 맥락에 따라 달라지는 '정답 변화'의 예. 같은 철도 이미지 쌍에서도 날씨 관점에서는 눈과 구름의 변화가, 철도 안전 관점에서는 선로에 나타난 열차가 핵심 변화가 된다.

그러나 기존 평가용 데이터는 특정 장면이나 변화에 집중돼 있거나, 각 상황에서 어떤 변화를 중요하게 봐야 하는지가 명확하게 정해져 있지 않은 경우가 많았다. 이 때문에 특정 데이터에서 좋은 성능을 보인 AI가 다양한 현실 상황에서도 필요한 변화를 제대로 찾아내는지 확인하기 어려웠다.

연구팀은 이러한 한계를 보완하기 위해 ▲자연 장면 ▲위성·항공 등 원격탐사 영상 ▲이미지 편집 ▲이상 탐지 등 4개 분야에서 51개의 현실적인 변화 상황을 선정하고, 사람이 변화 내용을 직접 설명한 4,996개의 이미지 쌍으로 C3-Bench를 구성했다.



▲ C3-Bench 개요. 자연 장면, 원격탐사 영상, 이미지 편집, 이상 탐지의 4개 분야

아래 도로·건설·재난·자원 개발·제품 결함 등 51개 변화 맥락을 구성하고, 각 맥락에 맞는 변화 설명 능력을 종합적으로 평가한다.

각 상황마다 AI가 '찾아야 할 변화'와 '무시해야 할 차이', '변화를 설명할 때 지켜야 할 기준'을 구체적으로 정했다.

이를 통해 같은 이미지 쌍이라도 눈에 보이는 차이를 단순히 나열하는 것이 아니라, 주어진 상황에서 중요한 변화를 제대로 골라 설명했지 평가할 수 있도록 했다.

AI가 내놓은 답을 평가하는 방식도 기존과 달랐다. 정답 문장과 같은 단어가 얼마나 포함됐는지만 비교하는 대신, ▲내용이 정확한지 ▲변화를 구체적으로 설명했는지 ▲문장이 자연스러운지 ▲주어진 상황에 맞는 내용인지를 함께 살펴 답변의 의미를 종합적으로 평가했다.

새 평가 방식이 사람의 판단을 얼마나 잘 반영하는지도 확인했다. 같은 AI 답변을 사람이 직접 평가한 결과와 여러 자동 평가 방식의 결과를 비교한 결과, 연구팀이 제안한 평가 방식이 기존 자동 평가 방식보다 사람의 판단과 더 가까운 것으로 나타났다.

연구팀은 이어 400개의 이미지 쌍을 대상으로 사람과 GPT-5.2가 각각 이미지 속 변화를 설명하도록 하고, 답변의 정확성·구체성·자연스러움·상황 관련성을 비교했다.

그 결과 4개 항목을 종합한 10점 만점 점수는 사람이 7.45점, GPT-5.2가 5.72점으로 1.73점의 차이를 보였다. 반면 문장의 자연스러움만 놓고 보면 GPT-5.2가 8.53점으로 사람(8.32점)보다 오히려 높았다.

이는 AI가 사람이 쓴 글에 가까운 자연스러운 문장을 만드는 능력과 실제 이미지에서 중요한 변화를 정확하게 찾아내는 능력이 반드시 같은 것은 아님을 보여준다.

연구팀은 또 이미지의 순서를 바꿨을 때도 같은 변화를 앞뒤에 맞게 설명하는지를 확인했다. 예를 들어 첫 번째 사진에서 "물체가 새로 생겼다"고 설명했다면, 사진의 순서를 뒤집었을 때는 같은 물체가 "사라졌다"고 설명해야 한다.

연구팀은 이러한 능력을 '가역성(Reversibility)'으로 평가했다. 그 결과 사람은 93%의 가역성 점수를 보인 반면 GPT-5.2는 61%에 그쳤다. 이는 최신 AI가 자연스러운 문장을 만드는 데는 뛰어나지만, 사진 속 변화의 방향까지 일관되게 이해하는 데는 여전히 한계가 있음을 보여준다.

AI가 사진 속 변화를 이해하는 과정에서 어떤 부분에서 주로 오류를 일으키는지도

분석했다.

그 결과 사진 속 물체를 제대로 알아보지 못하거나 종류·특성을 잘못 파악하는 시각적 인식 오류가 전체 오류의 63.3%로 가장 많았다. 이 밖에도 실제로 중요하지 않은 차이를 변화로 판단하거나, 변화의 의미 또는 위치 관계를 잘못 설명하는 오류도 확인됐다.

Perceptual Error	Spatial Error	Inconsistency Error	Change Collapse Error
			
GT: The tree in the center have been scorched, and a villa in the upper right corner has been destroyed. Prediction: There is no identifiable change.	GT: The yellow stool added beside the rectangular white desk. Prediction: The white rectangular table has been replaced by a round table, and a yellow stool added.	Prediction: Remove the three dogs from the left side and replace the field on the left with a small pond. Prediction (reverse): Add three additional dogs in the grassy field.	Prediction: A new surface crack has developed near the right edge of the pill. Prediction (reverse): There is no identifiable change.

▲ 최신 AI(GPT-5.2)의 대표 오류 사례. 왼쪽부터 물체나 속성을 놓치는 지각 오류, 위치 관계를 잘못 이해하는 공간 오류, 이미지 순서를 뒤집었을 때 설명이 서로 맞지 않는 불일치 오류, 한 방향에서만 변화를 인식하는 변화 붕괴 오류를 보여준다. 빨간 글씨는 잘못된 설명 부분이다.

김의환 교수는 "이번 연구는 AI가 '자연스럽게 설명하는 능력'과 '상황에 맞는 변화를 정확하게 이해하는 능력'이 반드시 같지는 않다는 점을 체계적으로 보여줬다"며 "C3-Bench가 향후 ▲재난·사고 현장의 변화 감시 ▲위성·항공 영상을 활용한 원격 탐사 ▲자율주행 ▲이상 탐지 등 다양한 현실 환경에서 AI의 변화 이해 능력을 평가하고 개선하는 연구의 공통 기반으로 활용되기를 기대한다"고 말했다.

AI학과 김의환 교수가 지도하고 김재우 석·박사통합과정생이 제1저자로 수행한 이번 연구는 과학기술정보통신부·한국연구재단 우수신진연구자지원사업, 정보통신기획평가원(IITP) 문제가설과 자가지도 기반의 자기주도 시각지능 기술 개발 사업, 과학기술정보통신부 지스트-이노코어(GIST-InnoCORE) 사업 등의 지원을 받았다.

연구 결과는 2026년 9월 8일부터 12일까지 스웨덴 말뫼에서 열리는 국제학술대회 'European Conference on Computer Vision(ECCV) 2026'에 채택됐다. ECCV는 유럽 컴퓨터비전협회(ECVA)가 주관하는 컴퓨터비전·머신러닝 분야의 주요 국제학술대회다.

한편 GIST는 이번 연구 성과의 기술사업화 및 산업적 활용에 관심 있는 기·기관의 문의를 창업혁신진흥원 기술사업화센터(hgmoon@gist.ac.kr)를 통해 받고 있다고 밝혔다.

1. 논문명, 저자 정보

- 학회명 : European Conference on Computer Vision (ECCV) 2026
- 논문명 : C3-Bench: A Context-Aware Change Captioning Benchmark
- 저자 정보 : 김재우(제1저자, GIST AI학과 석·박사통합과정생), 김형범(제2저자, GIST AI학과 학사과정생), 김의환(교신저자, GIST AI학과 부교수)