Intent Resolution via Inference-time Saccades

Tech ID: 34797 / UC Case 2026-815-0

BACKGROUND

Visual Question Answering (VQA) represents a fundamental challenge at the intersection of Computer Vision and Natural Language Processing, requiring systems to understand both visual content and linguistic queries to generate appropriate responses. While recent Vision-Language Models (VLMs) have achieved impressive performance on standard VQA benchmarks they continue to struggle with a pervasive real-world challenge: referential ambiguity. When multiple objects in an image could plausibly satisfy a query, current VLMs lack the contextual grounding to identify the intended object. A solution lies in leveraging eye movement fixations, a natural human behavior, to resolve referential ambiguity in open-ended VQA. During natural viewing and questioning, fixations reliably precede verbal references by several hundred milliseconds, reflecting both planning and execution in speech production. By aligning what is said with where and when people look, a time-locked, user-aligned signal can be obtained that helps disambiguate referential intent in ambiguous VQA scenarios.

DESCRIPTION

Researchers at the University of California, Santa Barbara and the Army Research Laboratory have solved referential ambiguity in VQA with Intent Resolution via Inference-time Saccades (IRIS), a novel training-free approach that uses real-time eye-tracking data to resolve referential ambiguity in visual question answering by guiding AI models to the intended object in complex images. IRIS leverages natural human eye movement fixations during question formulation to disambiguate which object a user refers to in ambiguous visual questions. IRIS integrates three key components: real-time eye-tracking to capture overt visual attention patterns, speech recognition to identify question timing and content, and multimodal large language models to generate responses. By synchronizing gaze data with speech onset, this system provides context to VLMs at inference time, improving accuracy without requiring any model retraining or architectural modification. Tested on 500 unique image-question pairs, IRIS more than doubles the accuracy on ambiguous queries and maintains performance on clear questions. It is compatible with existing and future VLMs and includes a new benchmark dataset, real-time interactive protocol, and evaluation tools.

ADVANTAGES

- Training-free approach requiring no model retraining or parameter updates
- Instant compatibility with existing and future vision-language models
- Significantly improves accuracy on ambiguous visual questions – from 35.2% to 77.2%

CONTACT

Donna M. Cyr
cyr@tia.ucsb.edu
tel: .

INVENTORS

- Eckstein, Miguel P.
- Karmakar, Srijita
- Madinei, Parsa

OTHER INFORMATION

KEYWORDS

Visual Question Answering, Natural Language Processing, eye-tracking, AI, eyes, eye movement, visual attention, Vision-Language Model, imaging diagnostics, accessibility

CATEGORIZED AS

- **Medical**
  - Imaging
  - Research Tools

RELATED CASES

2026-815-0

- Utilizes naturally occurring human eye-tracking data, enhancing user intent understanding
- Low computational overhead and scalable software-only solution
- Supports a wide range of applications through integration with common eye-tracking hardware
- Cost-effective to deploy across large user bases and reduces support costs

APPLICATIONS

- AI-powered customer service systems that require accurate user intent interpretation
- Retail platforms enhancing product identification and support via visual queries
- Healthcare imaging diagnostics with improved patient question understanding
- Education tools that interactively interpret student queries about visual content
- AR/VR experiences using integrated eye-tracking for context-aware assistance
- Automotive driver monitoring systems improving interface responsiveness to driver intent
- Accessibility devices enabling more precise communication through gaze signatures

ADDITIONAL TECHNOLOGIES BY THESE INVENTORS

- Optimal Perception and Eye-movement Response Assessment

University of California, Santa Barbara
Office of Technology & Industry Alliances
342 Lagoon Road, ,Santa Barbara,CA 93106-2055 |
https://www.tia.ucsb.edu
Tel: 805-893-2073 | Fax: 805.893.5236 | padilla@tia.ucsb.edu