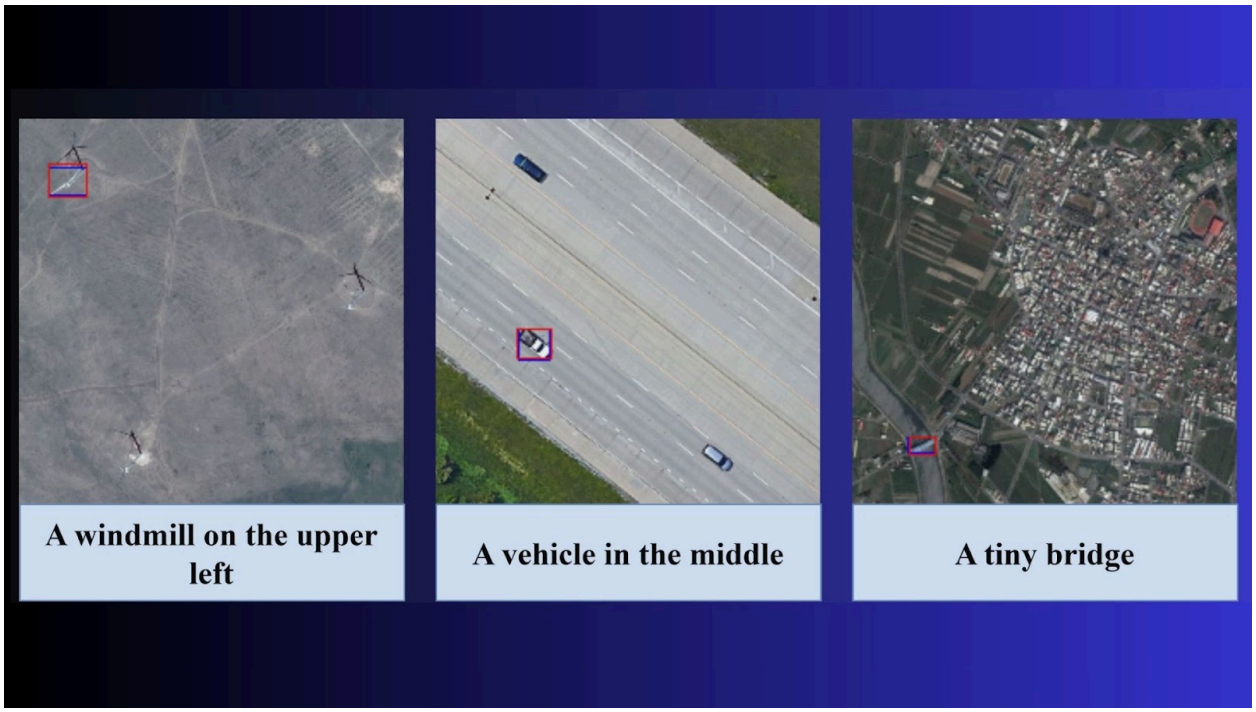
भुचित्रों का आकलन करके भाषा निर्देशों पर कार्य करता नया मॉडल

आईआईटी मुंबई के शोधकर्ताओं ने AMVG नामक एक ऐसा मॉडल निर्मित किया है, जो उपग्रह से प्राप्त या दूर संवेदी प्रतिमाओं के मशीनी विश्लेषण को मनुष्य के सामान्य भाषा निर्देशों के आधार पर खोज सकता है।



AMVG मॉडेल द्वारा खोजी गयी विषयवस्तु चौकोर में दर्शाती कुछ दूर संवेदी प्रतिमाएं एवं प्रतिमा के नीचे इस खोज हेतु दिए गए सामान्य भाषा निर्देश. (१) बायें की ओर ऊपरी भाग में पवनचक्की (२) प्रतिमा के मध्य में उपस्थित वाहन (३) छोटा पुल

श्रेय: अध्ययन के लेखक

अपने घर में एक छायाचित्र लें एवं कृत्रिम प्रज्ञा (आर्टिफिशियल इंटेलिजेंस, एआई) आधारित साधन से इसमें से बिल्ली अथवा टीवी रिमोट की खोज करने को कहें। आप शायद चकित रह जाएंगे की कितनी सटीकता से एआई तंत्र वस्तुओं को छायाचित्र में चिन्हांकित कर पाता है। किंतु उपग्रहों या ड्रोन से प्राप्त जटिल छवियों की दशा में प्राकृतिक भाषा संकेतों पर आधारित प्रश्न (प्रॉम्प्ट) के माध्यम से वस्तु की पहचान करने में अधिकांश अत्याधुनिक मॉडल अभी तक असफल रहे हैं।

प्रा. बिप्लब बनर्जी के नेतृत्व में भारतीय प्रौद्योगिकी संस्थान, मुंबई (आईआईटी मुंबई) के शोधकर्ताओं ने एक नवीन मॉडल निर्मित किया है, जो प्राकृतिक भाषा (नैचरल लैंग्वेज) संकेतों के माध्यम से दूर संवेदी छवियों (रिमोट सेंसिंग इमेजेस) में सन्निहित वस्तु की पहचान करने में सक्षम होगा। 'अडाप्टिव मोडेलिटी गाइडेड विजुअल ग्राउंडिंग' (AMVG) नामक यह मॉडल छवि में उपस्थित वस्तुओं की पहचान करने के साथ-साथ उपयोगकर्ता के द्वारा पूछे गए प्रश्नों को भलीभांति समझ सकता है, भले ही वे प्रश्न अस्पष्ट अथवा संदर्भ विशिष्ट क्यों न हों।

उदाहरण स्वरूप यदि हम 'बाढ़ग्रस्त नदी के निकट स्थित समस्त क्षतिग्रस्त भवनों को खोजो' यह निर्देश देते हैं तो एक मनुष्य विश्वसनीय रूप से इसका निष्पादन कर सकता है। किंतु अनेक छोटी छोटी वस्तुएं जिसमें उपस्थित हो ऐसी सैकड़ो छवियों को मनुष्य की भांति या उससे भी अधिक दक्षता के साथ मशीन द्वारा मिनिट भर में परखने (स्कैन करने) हेतु मशीन को प्रशिक्षित करना अत्यंत महत्वपूर्ण है। एक कंप्यूटर अथवा यंत्र को दैनिक वार्तालाप की भाषा में दिये गए निर्देशों का बोध करने हेतु प्रशिक्षित करना एवं छवियों में सन्निहित विवरण के साथ इनका मेल होना विजुअल ग्राउंडिंग या फ्रेस ग्राउंडिंग (visual or phase grounding) कहलाता है।

"*दूर संवेदी छवियाँ अधिक विवरण के साथ समृद्ध होती हैं, किंतु स्वचालित रीति से इसकी विवेचना कर पाना चुनौतीपूर्ण कार्य है। यद्यपि विजुअल ग्राउंडिंग में उन्नति हो चुकी है, तब भी वर्तमान मॉडल दूर संवेदी छवियों का भलीभांति आकलन कर पाने में असफल रहे हैं, विशेषकर जब मनुष्य से दिये गए निर्देश अस्पष्ट अथवा विशिष्ट संदर्भ पर निर्भर होते हैं*," इस अध्ययन की प्रमुख लेखिका एवं पीएचडी शोधार्थी शबनम चौधुरी का कहना है।

वर्ष-प्रतिवर्ष दूर संवेदी डेटा तीव्रता से निरंतर बढ़ रहा है। पृथ्वी के ऊपर अधिकाधिक दूरियों से (उपग्रह, ड्रोन या वायुयान के द्वारा) ली गई इन छवियों में वातावरणीय कोलाहल (नॉइज़) एवं मापन विविधता (स्केल वेरिएशन) के साथ अनेक विषयवस्तु (ऑब्जेक्ट) सूक्ष्म रूप में निहित होती हैं। इन छवियों में एक भवन एक विमानपथ के रूप में तथा विमानपथ एक नदी के रूप में दिखाई पड़ सकता है। यह अध्ययन बताता है कि किस प्रकार से एक उत्तम अनुवादक के रूप में कार्य करते हुए AMVG, मानवीय दैनिक वार्तालाप की भाषा में प्राप्त निर्देशों की विवेचना करता है एवं विषयवस्तु को विश्वसनीय रूप से पहचान पाता है। आईआईटी मुंबई का यह अध्ययन *ISPRS की 'फोटोग्रामेट्री एंड रिमोट सेंसिंग'* शोध पत्रिका में प्रकाशित किया गया है।

देखते हैं आईआईटी मुंबई के शोधकर्ताओं ने यह असंभव सा प्रतीत होने वाला कार्य कैसे किया। चौधुरी बताती हैं कि विजुअल ग्राउंडिंग हेतु अधिकांश मॉडल आज द्वि-चरणीय पद्धति का उपयोग करते हैं: प्रथम वे छवि के विभिन्न भागों का प्रस्ताव करते हैं, तत्पश्चात उनको श्रेणीबद्ध करते हैं। दूसरी ओर AMVG चार प्रकार के अभिनव घटकों से युक्त है: मल्टी-मोडल डिफॉर्मेबल अटेंशन लेयर, मल्टी-स्टेज टोकेनाइज़्ड एनकोडर (MTE), मल्टी-मोडल कंडीशनल डिकोडर, एवं अटेंशन अलाइनमेंट लॉस (AAL) नामक चार अभिनव स्तर इसके भाग हैं।

AMVG का प्राथमिक स्तर, छवियों के प्रत्येक पिक्सेल (डिजिटल छवि का सबसे छोटा भाग) का विश्लेषण करने के स्थान पर, पूछे गए प्रश्नों से संबंधित छवि के विशिष्ट भागों की प्राथमिकता सुनिश्चित करता है। द्वितीय स्तर MTE कुशल परिशोधन (रिफाइन) एवं आकलन का कार्य करता है, जो कठिन निर्देशों को एक-एक कर समझते हुए छवि की विशेषताओं एवं विवरणों में संबंध जोड़ता है। तीसरा मल्टी-मोडल कन्डीशनल डिकोडर, एकमात्र अनुमान लगाने के स्थान पर मॉडल द्वारा विषयवस्तु की खोज को क्रमशः परिष्कृत करता है। यह ठीक वैसा है जैसे कि एक जासूस कुछ संभावनाओं को हटा कर खोज में संदिग्धों की संख्या को कम करता है। मॉडल का चौथा स्तर एवं सबसे विलक्षण कार्य एक नवीन प्रशिक्षण तकनीक है, जिसे अटेंशन अलाइनमेंट लॉस (AAL) कहते हैं। यह अभिनव प्रशिक्षण पद्धति इस प्रकार कार्य करती है जैसे एक शिक्षक अपने छात्र के ध्यान को अपेक्षित स्थान पर केंद्रित करने हेतु मार्गदर्शन कर रहा हो।

"AAL को एक प्रशिक्षक के रूप में देखा जा सकता है। जब एक मनुष्य कहता है 'देखो उस ट्रक के पीछे एक कार खड़ी है' तब उसकी आँखें भलीभांति जानती हैं कि परिदृश्य में किस स्थान पर ध्यान केंद्रित करना

है। किन्तु AMVG एक मशीन होने के कारण उसे इस प्रकार के सहज ज्ञान तथा आकलन के लिए सहायता की आवश्यकता होती है। और AAL ठीक यही कार्य करता है। यह AMVG मॉडल को बताता है कि खोज के लिए ध्यान कहा केंद्रित करना है। यदि मॉडल का "ध्यान" एक सीमा से अधिक दूर तक विचलित होता है तो AAL इसे हलकेसे उचित स्थान की ओर धकेल देता है," चौधुरी स्पष्ट करती हैं। AMVG के ये चार स्तरों के घटक एक साथ मिलकर AMVG मॉडल को "संदर्भ के साथ देख पाने" एवं "बारीकी से निर्देश सुन पाने" की ऐसी क्षमता प्रदान करते हैं कि पूर्व में किये गए अन्य प्रयासों से यह मॉडल बहुत अग्रणी हो जाता है।

यह मात्र तकनीकीजन्य उन्नति नहीं है, अपितु इस मॉडल के अनेक प्रत्यक्ष उपयोग है। इनमें नागरी योजना, कृषि उत्पादकता, आपदा प्रतिक्रिया एवं सैन्य संनिरीक्षण (सर्वेलंस) सम्मिलित है।

"आपदा प्रबंधन इसके सबसे उत्साहजनक अनुप्रयोगों में से एक है," चौधुरी बताती हैं। बाढ़, भूकंप, या दावानल (वाइल्ड फायर) जैसी आपदाओं के समय, उदाहरण स्वरूप इस मॉडल से यह पूछा जा सकता है कि 'राजमार्ग के निकट के क्षतिग्रस्त भवनों को दिखाएं।' इस प्रकार उस स्थान के सटीक निर्देशांक प्राप्त किये जा सकते हैं। इसी प्रकार 'सीमावर्ती घने क्षेत्रों में छद्मावरण (कॅमफ्लॉज) वाहनों" को खोजने का प्रयास करते सैनिक हों या 'सिंचाई क्षेत्रों के निकट स्थित उपज में पीले भागों' को खोजते किसान, यह मॉडल वास्तविक समय (रिअल-टाइम) में जानकारी प्रदान कर सकता है।

इस मॉडल की एक विशेषता यह भी है कि शोधकर्ताओं ने इसे पूर्णतः मुक्त स्रोत (ओपन सोर्स) रखते हुए AMVG का संपूर्ण कार्यान्वयन GitHub पर सार्वजनिक रूप से उपलब्ध कराया है। दूर संवेदी शोध कार्यों के क्षेत्र में यह एक दुर्लभ उपलब्धि है।

"प्राकृतिक छवियों के विजुअल ग्राउंडिंग पर कार्य कर रहे समुदाय में मुक्त स्रोत व्यवस्थाएं होती हैं, तथापि दूर संवेदी छवियों के क्षेत्र में यह दुर्लभ ही है। अधिकांश अत्याधुनिक रिमोट सेंसिंग मॉडल मुक्त नहीं हैं या केवल आंशिक रूप से ही मुक्त किये गए हैं। इससे सामूहिक प्रगति तीव्र गति से नहीं हो पाती," चौधुरी बताती हैं।

"AMVG को मुक्त स्रोत रखना एक विचारपूर्वक तथा उतनी ही गहनता से व्यक्तिगत निर्णय भी है। हमारा विश्वास है कि वैज्ञानिक प्रभाव वास्तविकता में तब होता है जब आपके कार्य में उपयोगकर्ताओं के लिए किसी शुल्क की बाध्यता नहीं होती। हमें विश्वास है कि अपने इस मॉडल को पूर्ण रूप से प्रकाशित करके हम पारदर्शिता, पुनरुत्पादकता एवं दूर संवेदी विजुअल ग्राउंडिंग शोध के क्षेत्र में तेज प्रगति को प्रोत्साहित कर सकते हैं," वह कहती हैं।

यद्यपि कोई भी मॉडल आदर्श अथवा परिपूर्ण नहीं है। AMVG अभी भी उच्च कोटि के एवं टिप्पण सहित डेटा समूह (एनोटेटेड डेटसेट्स) पर निर्भर करता है। जब ऐसे प्रकार की छवियाँ अथवा संवेदक, जो मॉडल ने पूर्व में कभी नहीं देखे, मॉडल के सामने आते हैं तब मॉडल के प्रदर्शन में भिन्नता आ सकती है। यद्यपि पूर्व के अन्य मॉडलों की तुलना में यह मॉडल अधिक दक्ष है, तथापि वास्तविक समय में अथवा 'एज डिवाइसेज' (edge devices; डेटा स्त्रोत को नेटवर्क से जोड़ने वाले उपकरण जिनका 'इंटरनेट-ऑफ़-थिंग्स' में प्रयोग होता है) में उपयोग किये जाने हेतु इन्हें अभी और अनुकूलित किये जाने की आवश्यकता है।

किंतु अभी कार्ययोजना स्पष्ट है। शोधदल ने इसके सेंसर-अवेयर प्रकार (संवेदक की जानकारी रखने वाला प्रकार), संरचनात्मक विजुअल ग्राउंडिंग (उदाहरण हेतु 'वृक्ष के निकट स्थित नीले टैंक के पीछे छोटी सी

झोपड़ी') एवं लार्ज वीजन-लैंग्वेज मॉडल पर कार्य करना पूर्वसे ही आरंभ किया है ताकि इसे अन्य प्रकार के संवेदकों, भौगोलिक स्थितियों एवं अन्य कार्यों हेतु उपयोग में लाया जा सके।

"अंततः हम दूर संवेदन के ऐसे एकीकृत बोध से युक्त मॉडल को आगे लाना चाहते हैं जो अपने किसी भी स्वरूप में प्राकृतिक भाषाओं के साथ किसी भी छवि की ग्राउंडिंग में, विवरण प्रस्तुत करने में, खोज करने में एवं तर्क करने में सक्षम हो," चौधुरी कहती हैं।

आईआईटी मुंबई का AMVG मॉडल तकनीकी रूप से सक्षम होने के साथ-साथ बड़े स्तर पर किये जाने वाले वास्तविक अनुप्रयोगों में दूर संवेदन की उपलब्धता सुनिश्चित करता है। हम कैसे बात करते हैं एवं मशीन इसका विश्लेषण कैसे करती है, इस अंतर को पाटकर हम पृथ्वी का निरीक्षण करने वाले साधनों को उनकी पहुँच में ले आये हैं जिन्हें इसकी अधिक आवश्यकता है। यह हमारी सुनियोजित एवं सुव्यवस्थित संसार की ओर ले जानेवाली यात्रा सुलभ करता है।

VETTED / UNVETTED	Vetted
Title of Research Paper	Improving visual grounding in remote sensing images with adaptive modality guidance
DOI of the Research Paper as a link	https://doi.org/10.1016/j.isprsjprs.2025.02.031
List of all researchers with affiliations	Shabnam Choudhurya, Indian Institute of Technology, Bombay, India; Pratham Kurkure, Indian Institute of Technology, Bombay, India; Biplab Banerjee, Indian Institute of Technology, Bombay, India, and Centre for Machine Intelligence and Data Science (C-MInDS), Bombay, India
Email of researcher/s	Shabnam Choudhury <choudhury.shabnam6@gmail.com> Biplab Banerjee <getbiplab@gmail.com>
Writer name	Deekshith Pinto
Transcreator name	Somnath Danayak
Credits to Graphic:	Authors of the study
Subject [FOR EDITOR] - Please Highlight in RED (Multiple allowed)	Technology

VETTED / UNVETTED	Vetted
Article to be Sectioned Under [FOR EDITOR] - Please Highlight in RED	Deep Dive
Social Media TAGS separated by Comma	
Social Media Posts Suggestions/ Links to interesting relevant content [optional] [writer]	1. IIT Bombay researchers build a new model to bridge the gap between how humans prompt and how machines analyse satellite or remote sensing imagery. Details at <link> 2. IIT Bombay's newly proposed Adaptive Modality-guided Visual Grounding (AMVG) framework recognises what's in the image, and also understands what the user is asking about it, even with ambiguous or contextual prompts. Read on for more <link> 3. IIT Bombay researchers propose a new and technically superior Adaptive Modality-guided Visual Grounding (AMVG) framework, which can potentially make remote sensing truly accessible for large-scale, real-world applications. More about it at <link>
Social Media Handles to be added	@iitbombay
Social Media handles of writer	LinkedIn: Deekshith Pinto (https://www.linkedin.com/in/deekshith-pinto-editor/) Twitter: deekshith_np
Social Media handles of researchers	
Funding information (Source: Research paper)	NA
Conflict of Interest/Competing Interest information (Source: Research paper)	The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

VETTED / UNVETTED	Vetted
Co-PI information (Source: Research paper)	NA
Location:	Mumbai