How IISc Bangalore fine-tuned **7B** and **13B** open-weight vision models to read hand-drawn sketches

IISc Bangalore's Visual Computing Lab built O3SLM - an open-weight vision model that understands hand-drawn sketches, released in 7B and 13B versions, accepted at AAAI 2026, and the first of its kind in the world.

NEYSA SETUP

Service
Velocis Bare Metal

Model
LLaVA-v1.5

OUTCOME

World-first

Open-weight sketch vision model

Accepted at AAAI 2026

State-of-the-art on four sketch benchmarks

OVERVIEW

Hand-drawn sketches have always been a blind spot for AI. Commercial vision models are trained on photographs, so while they can describe a picture in detail, they cannot make sense of a rough sketch drawn by hand.

IISc's Visual Computing Lab built O3SLM to close that gap. With no suitable training data available, the team generated 32 million sketches from scratch, then fine-tuned LLaVA-v1.5 on Neysa Velocis in two stages, first aligning sketches with images, then tuning the model to follow sketch-based instructions. Tested across four sketch datasets, O3SLM reached state-of-the-art performance.

CHALLENGE

> " When you are working on a problem nobody has solved before, the last thing you want is infrastructure telling you what is and is not possible. We needed the freedom to attempt the full research design without compromise. "

Anirban Chakraborty
Professor, Visual Computing Lab, Indian Institute of Science Bangalore

Teaching AI to understand sketches is a harder engineering problem than it looks. Three things had to come together for O3SLM to work.

No training data existed

Every major sketch dataset was too small or too narrow to train on. The IISc team had to generate millions of sketches automatically from photographs - a pipeline that had to run continuously before training could begin.

The model architecture was inherently memory-hungry

O3SLM processes sketches, photographs, and language in a single pass. Running it all together in training requires substantially more GPU memory than most shared GPU clusters can provide - any compromise would create a weaker model.

Research at this level requires constant iteration

Reaching state-of-the-art needs multiple training runs. IISc ran evaluations across four sketch datasets and repeated experiments to find what actually drove performance.

What the research demanded	Buying your own GPUs	General-purpose cloud	Required for production
Getting started	• Months of procurement, then someone has to run it	• Powerful, but you set up everything yourself	• Ready-to-use environment, up and running on day one
Freedom to iterate	• One shared machine, queued behind other projects	• Every repeat run bills again, by the hour	• Dedicated capacity to run experiments without waiting
Scaling with the research	• Fixed capacity, hard to grow once it's setup	• Scales, but cost climbs steeply as you do	• Easy to scale up with a predictable cost, and timeline

TRANSFORMATION

> " Neysa gave us the compute headroom to run this project the way it needed to be run. The dataset, the training pipeline, the evaluations - none of it required us to scale down our ambitions to match what was available. "

Anirban Chakraborty
Professor, Visual Computing Lab, Indian Institute of Science Bangalore

The IISc team ran every phase of O3SLM's development on Neysa Velocis. Starting with 33 million automatically generated sketch-image pairs, they fine-tuned the model in two sequential stages on 815,000 multimodal instruction pairs and evaluated it across four benchmark datasets to confirm state-of-the-art performance.

Compute capacity that matched the research design

Dedicated, high-memory GPU nodes gave the team the headroom to run O3SLM's full architecture in training without compromising on model size. Both the 7B and 13B versions deliver state-of-the-art results.

Sustained throughput for large-scale data generation

Neysa delivered the sustained compute to fine-tune O3SLM at full scale - processing sketches, photographs, and language in a single pass, with no cuts to model size or batch size.

The iteration speed that frontier research demands

Getting to state-of-the-art requires more than a single training run. The team ran evaluations across four sketch datasets and repeated experiments across design variations.

MEASURED OUTCOMES

O3SLM, fine-tuned on Velocis Bare Metal

Model release
7B & 13B
Open-weight sketch vision models, the first of their kind.

Peer recognition
AAAI 2026
Accepted at a first-class AI research conference.

Benchmark performance
State of the art
Beating GPT-4o and Gemini 1.5 Pro on sketch tasks.

Sketch retrieval accuracy
65 vs 11
Acc@1 on Sketchy, up from 11 for the base LLaVA model.

Sketches generated
32M
Instance-level sketches across 3.7M images, built from scratch.

Tasks unified
4 in one model
Localization, counting, retrieval and VQA from sketches.

Start your production AI deployment with Neysa

Get in touch with our experts.

Request a demo | hello@neysa.ai

neysa