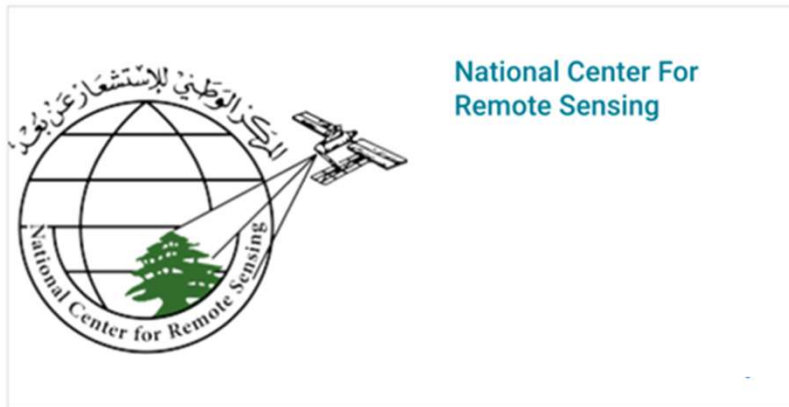# How SAM would revolutionize geospatial applications?
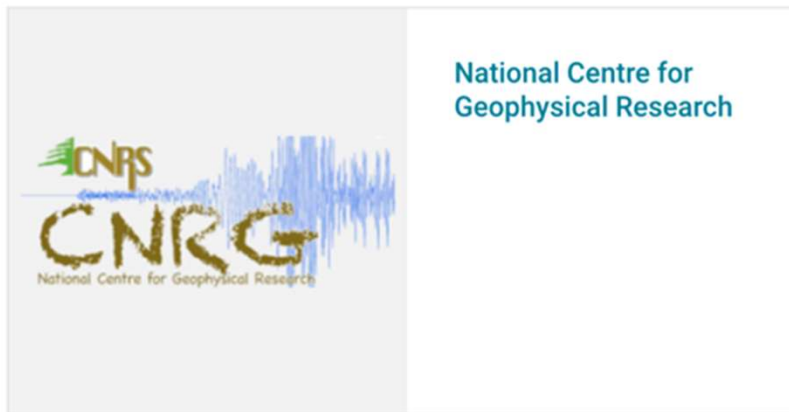
Ali Ghandour, PhD
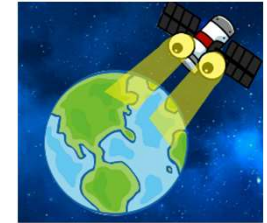
26 May 2023

CNRS-L Research Centers

National Centre for
Geophysical Research

National Centre for
Marine Sciences

National Center For
Remote Sensing

Lebanese Atomic
Energy Commission

GEOspatial Artificial Intelligence (GEOAI)
www.geogroup.ai
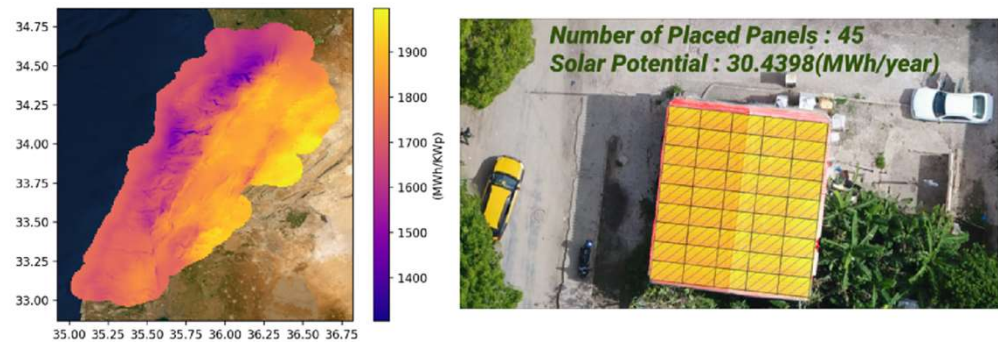

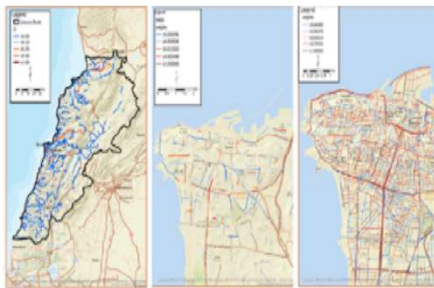**Urban-AI Mapping**

720K buildings

First Lebanese Urban Map
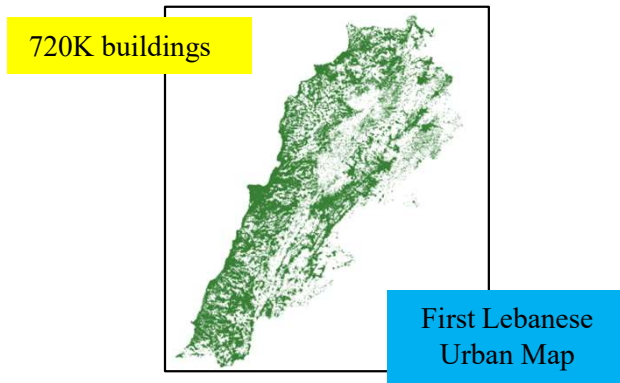
Solar rooftop potential map

Road Crashes Observatory

Water Body Monitoring

Crop Monitoring

The idea and practice of looking down at the Earth's surface emerged in the 1840s

- Cameras secured to tethered **balloons** for purposes of topographic mapping
- **Pigeon fleet** as a novel platform that operated in Europe
- Cameras mounted on **airplanes** for aerial large views (military reconnaissance)
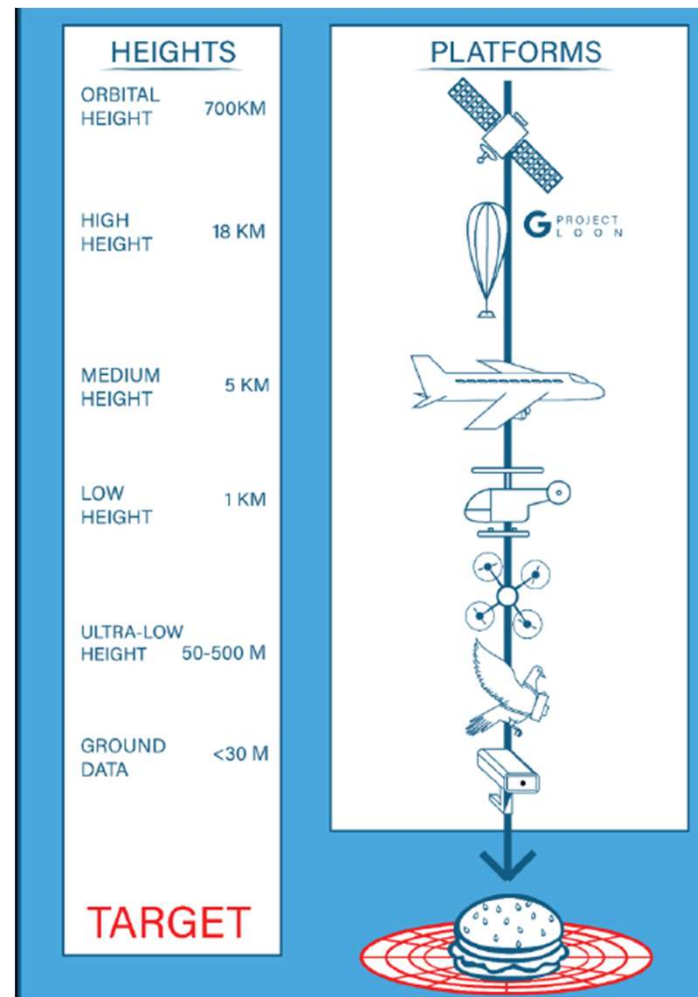

Balloons (1868)   Pigeons (1903)   Airplane (1909)

source: https://www.gabriele-cavallaro.com/teaching/parallel-and-scalable-machine-learning-for-remote-sensing-big-data-ws-2020-2021
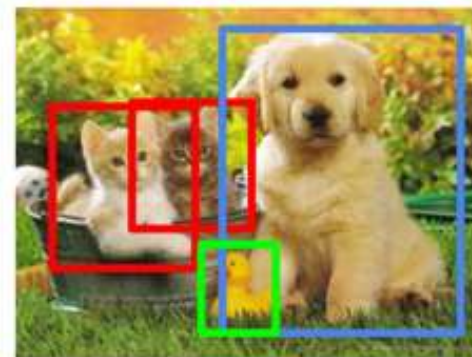
RS Platforms

source: https://www.gabriele-cavallaro.com/teaching/parallel-and-scalable-machine-learning-for-remote-sensing-big-data-ws-2020-2021

"How SAM would revolutionize geospatial applications?", Dr. Ali J. GHANDOUR, 26 May 2023

■ **Classification:**

source: https://www.geeksforgeeks.org/object-detection-vs-object-recognition-vs-image-segmentation/
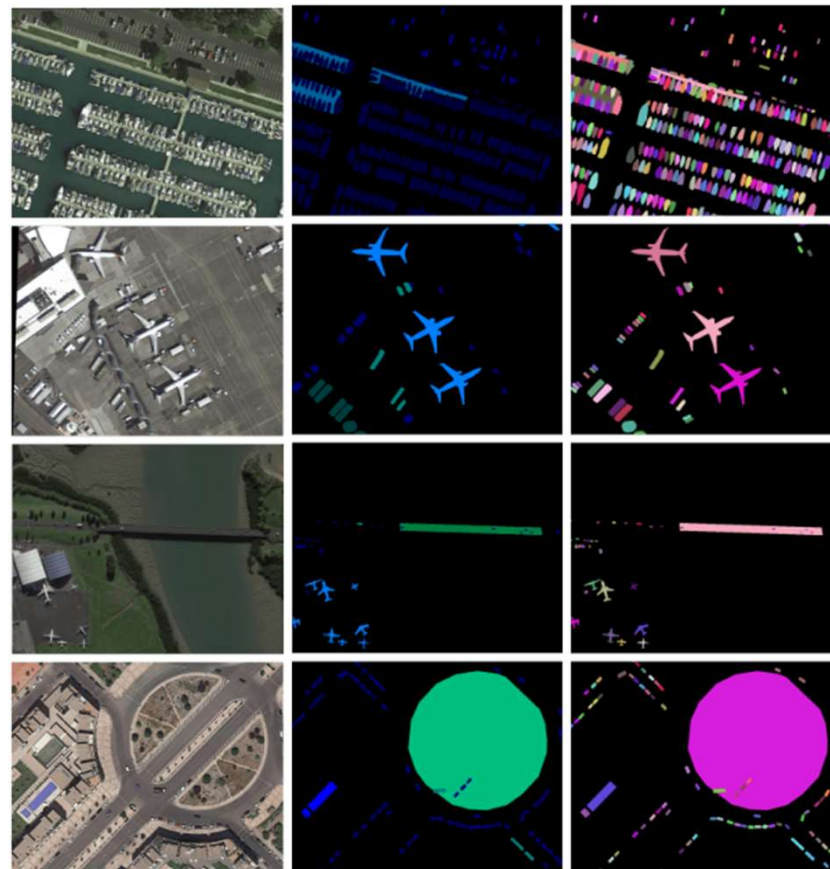
■ **Detection:**

CAT, DOG, DUCK

■ (Semantic) Segmentation:

Source: https://towardsdatascience.com/semantic-segmentation-of-aerial-imagery-using-u-net-in-python-552705238514
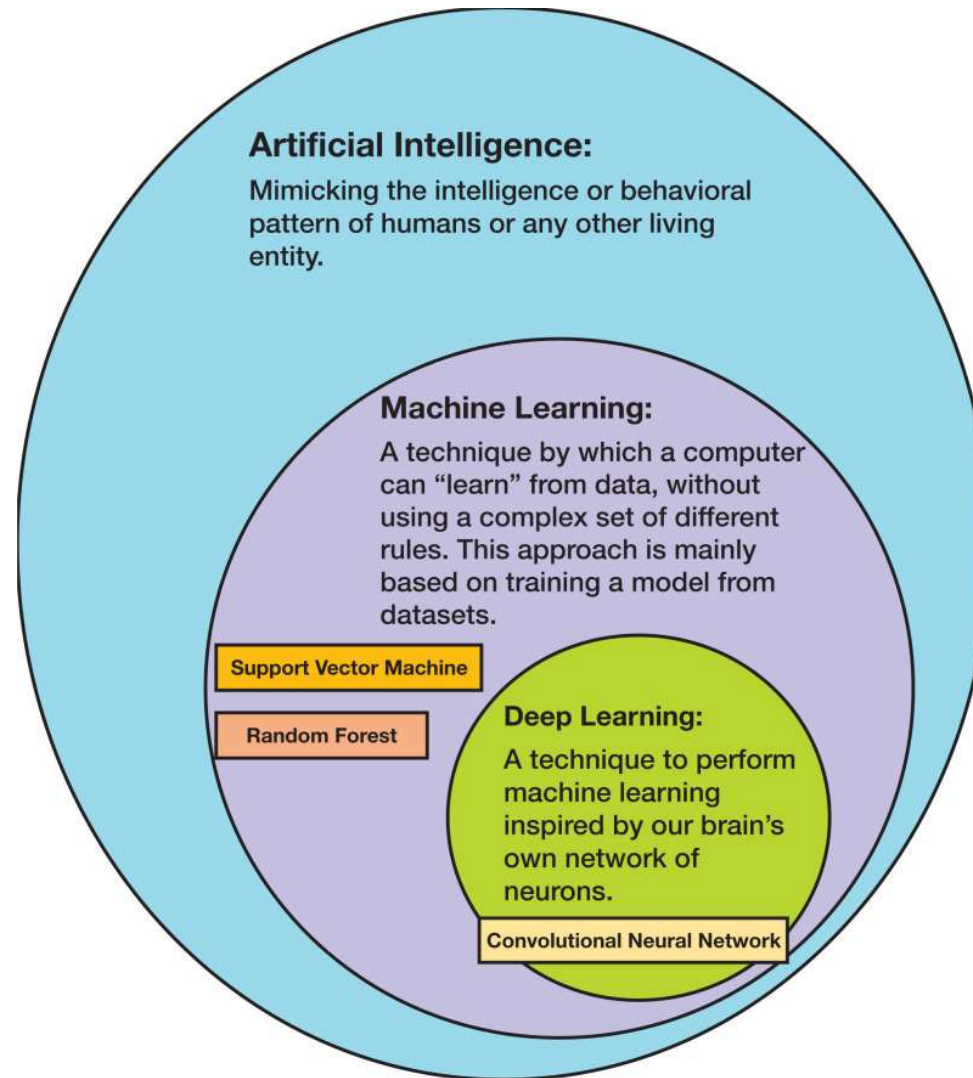
- ## (Instance) Segmentation:

(a) Original image   (b) SS maps   (c) IS maps

Source: https://arxiv.org/pdf/1905.12886.pdf

AI

Artificial Intelligence:
Mimicking the intelligence or behavioral pattern of humans or any other living entity.

Machine Learning:
A technique by which a computer can "learn" from data, without using a complex set of different rules. This approach is mainly based on training a model from datasets.

Support Vector Machine

Random Forest

Deep Learning:
A technique to perform machine learning inspired by our brain's own network of neurons.

Convolutional Neural Network

Source: https://www.ncbi.nlm.nih.gov/books/NBK570711/

Deep Learning

- **model extracts the feature set by itself without supervision (XAI field):**
  - ☐ faster
  - ☐ more accurate.

- **Training set**
- **Validation set**
- **Test set**

Deep Learning – *cont.*

- model extracts the feature set by itself without supervision (XAI field):
  - ❑ faster
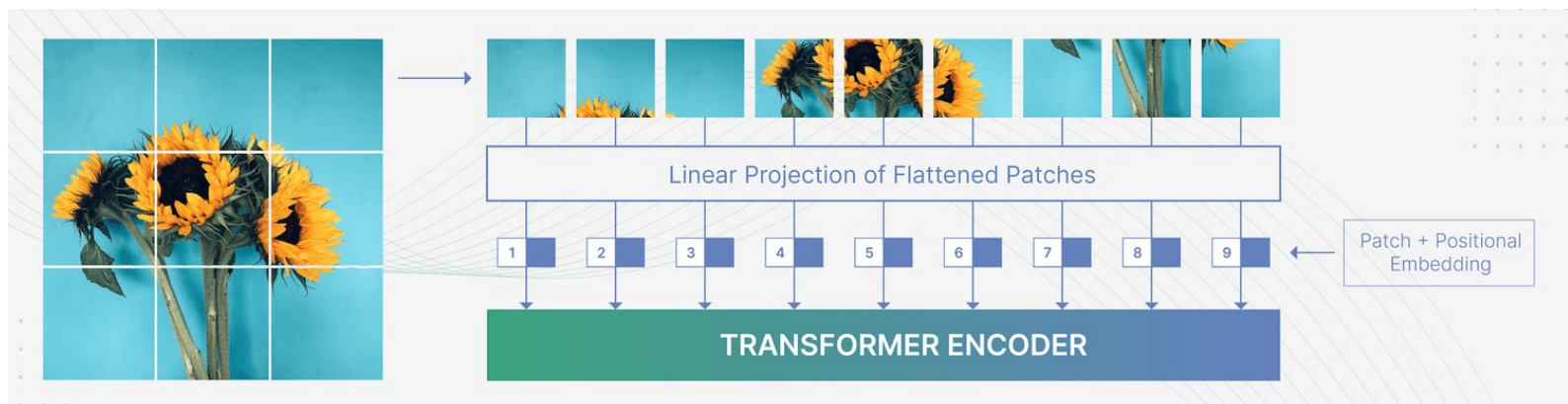 - ❑ more accurate.

- Training set
- Validation set
- Test set

Deep Learning – *cont.*

- ## CNN:

Source: https://viso.ai/deep-learning/vision-transformer-vit/

- ## Transformers:

Source: https://www.picsellia.com/post/are-transformers-replacing-cnns-in-object-detection/

OOD Generalization

- **Out-of-Distribution Generalization:**
  - ❑ testing distribution is unknown and different from the training

source: https://www.gabriele-cavallaro.com/teaching/parallel-and-scalable-machine-learning-for-remote-sensing-big-data-ws-2020-2021

Foundation Models

- **coined in July 2022:**
  - ❑ https://arxiv.org/pdf/2108.07258.pdf
  - ❑ Large (deep neural network) model
  - ❑ trained on broad data (generally using self-supervision at scale)
  - ❑ usually multimodal
  - ❑ can be adapted to a wide range of downstream tasks

- ■ **transformer model on a large scale.**

- ■ **widely used in NLP: Large Language Models (LLMs)**

- ■ **human brain:**
  - ❏ **~86 billion neurons**
  - ❏ **100 trillion synapses**

Source: https://huggingface.co/blog/large-language-models/

SAM
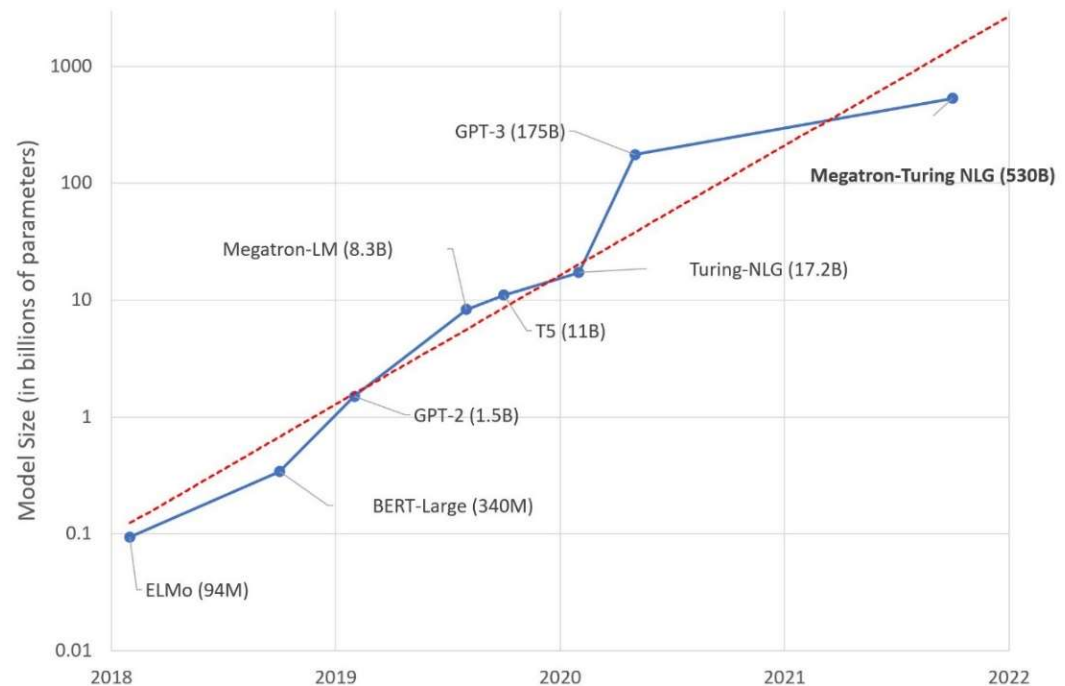
- **new model from Meta AI: April 2023**
  - ❏ https://arxiv.org/pdf/2304.02643.pdf

- **effectively segment any object in any image, without the need for additional task-specific or dataset-specific training**

- **various input prompts:**
  - ❑ points
  - ❑ bounding boxes
  - ❑ masks
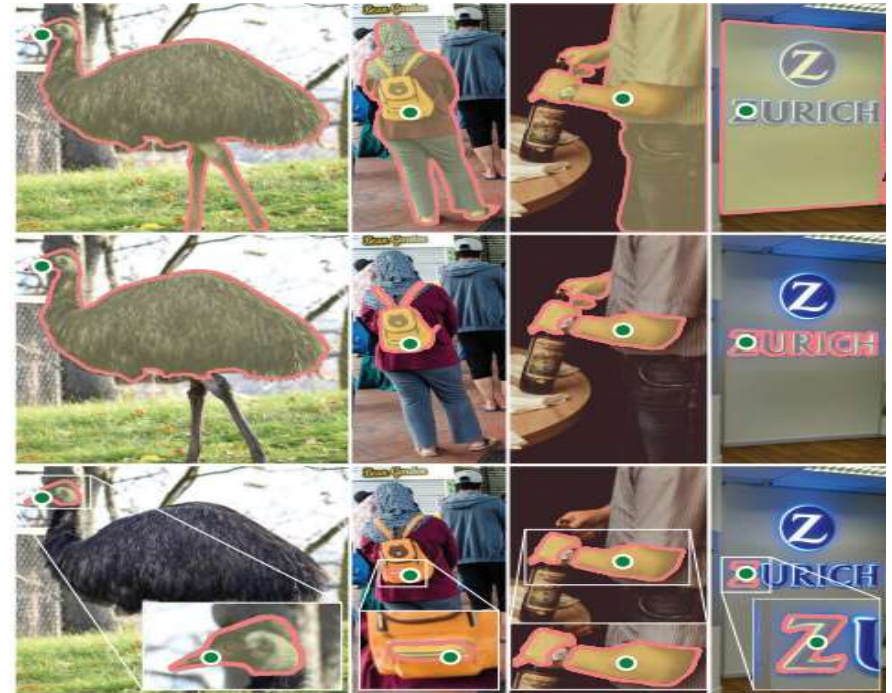
- **Ambiguity-Aware Design:**
  - ❑ handle ambiguous situations
  - ❑ multiple masks for a single prompt

SA-1B Dataset Overview

- **Dataset Composition:**
  - ❏ 11M diverse, high-res images (3300x4950 pixels)
  - ❏ 1.1B segmentation masks (99.1% auto-generated)
  - ❏ High-quality masks
  - ❏ Significantly larger than any segmentation datasets

Figure 10: Zero-shot edge prediction on BSDS500. SAM was not trained to predict edge maps nor did it have access to BSDS images or annotations during training.
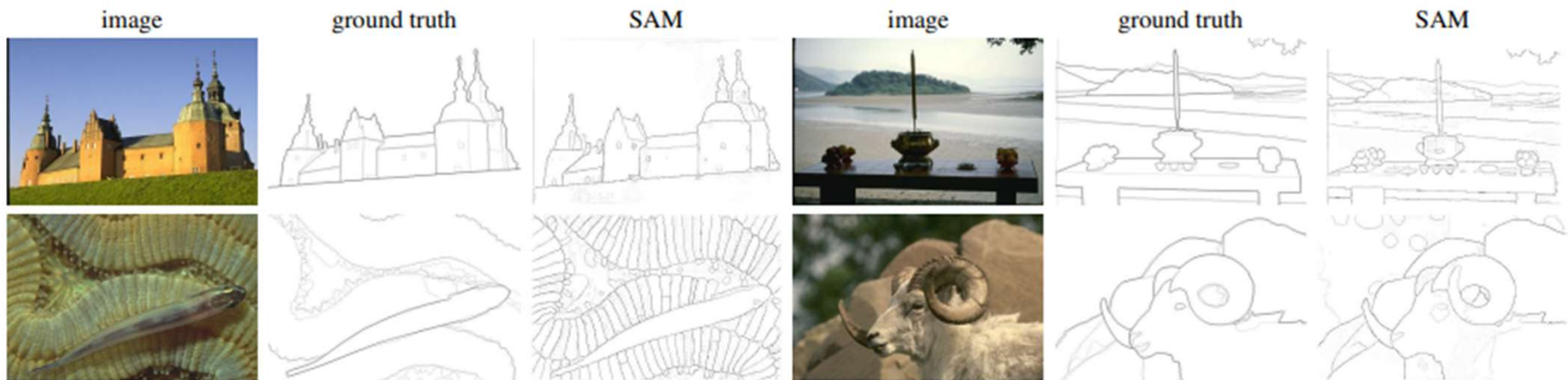
Figure 15: Additional visualizations of zero-shot edge predictions on BSDS500. Recall that SAM was not trained to predict edge maps and did not have access to BSDS images and annotations during training.

Zero-Shot Instance Segmentation
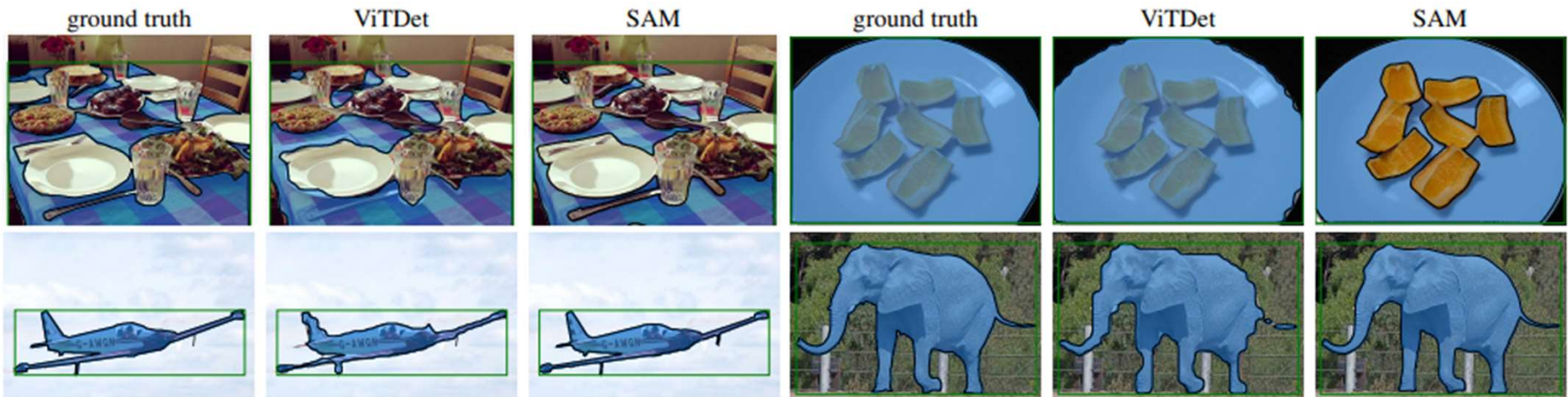
Figure 16: Zero-shot instance segmentation on LVIS v1. SAM produces higher quality masks than ViTDet. As a zero-shot model, SAM does not have the opportunity to learn specific training data biases; see top-right as an example where SAM makes a modal prediction, whereas the ground truth in LVIS is amodal given that mask annotations in LVIS have no holes.

Zero-Shot Text-to-Mask
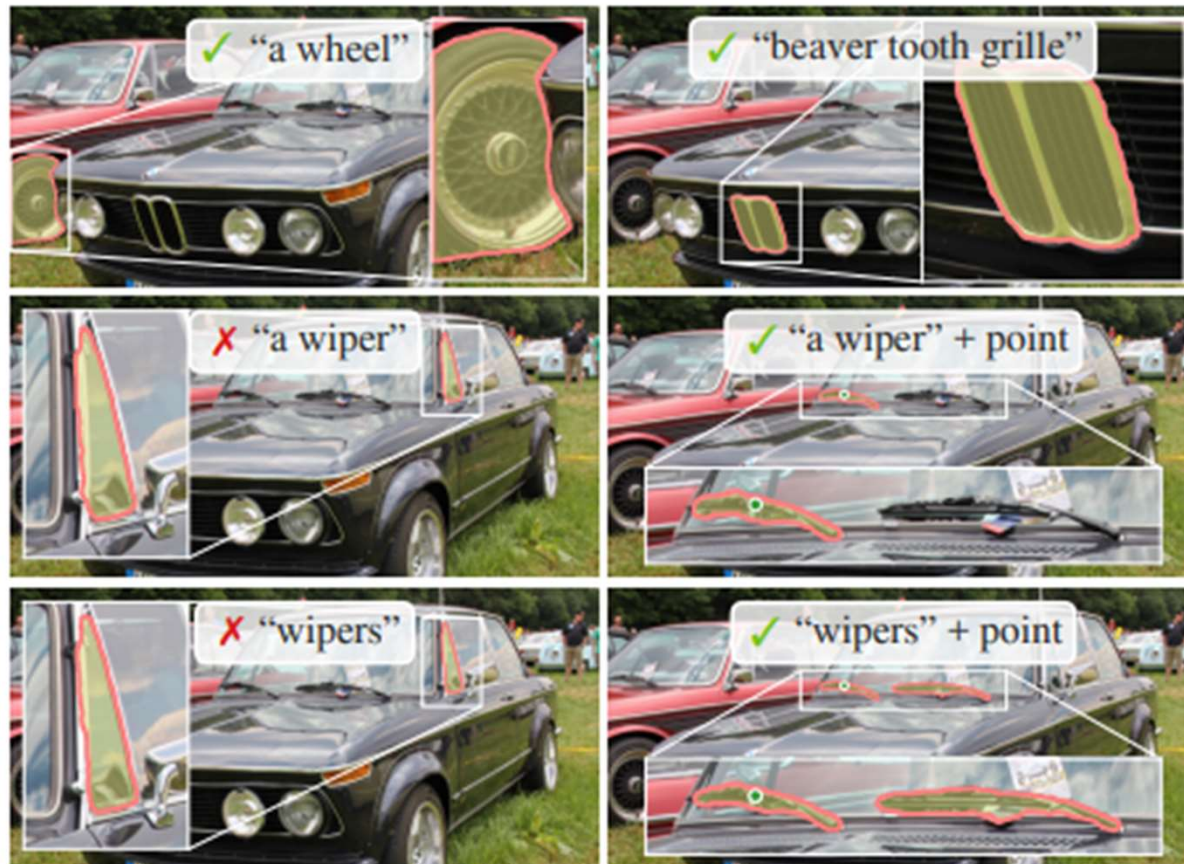
Figure 12: Zero-shot text-to-mask. SAM can work with simple and nuanced text prompts. When SAM fails to make a correct prediction, an additional point prompt can help.

Adaptation

Annotation Business

Descartes Labs
@DescartesLabs

Video of the Descartes Labs technology incorporating the Segment Anything Model (SAM), showing fast prototyping leveraging SAM powered by the Descartes Labs platform to retrieve imagery & display vectorized output in an interactive way.

More Adaptation
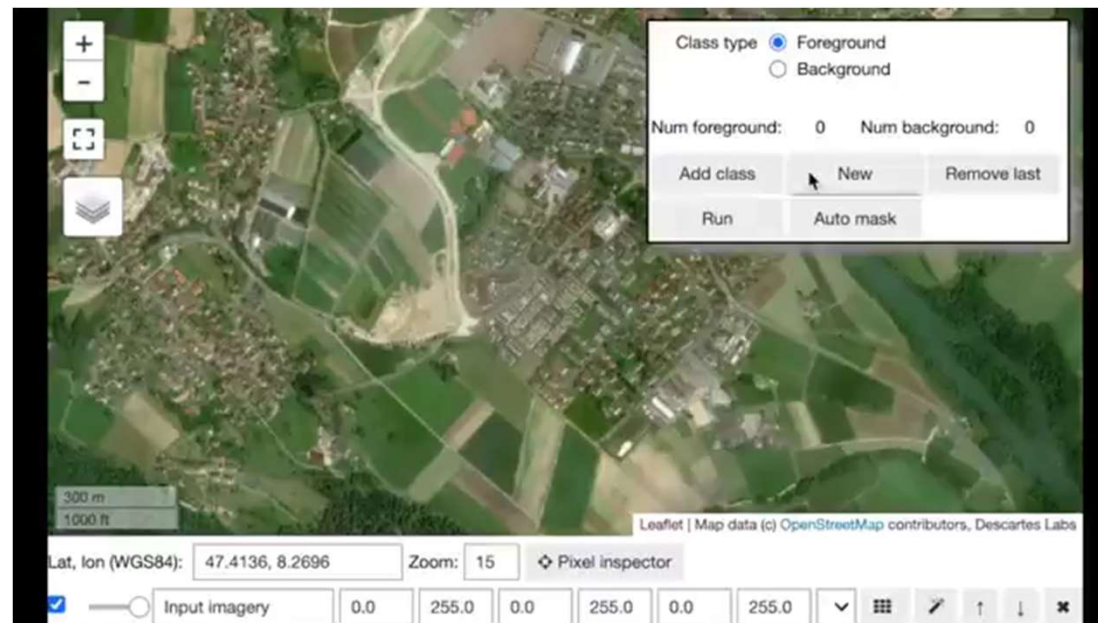
- **TrackSAM:**

Live Demo!

- **Ali: graduate student at GEOAI group**
- **Hasan: research engineer at GEOAI group**

Thank you!