One Prompt Fits All: Visual Prompt-Tuning for Remote Sensing Segmentation

Marshall Wang,[1] John Willes,[1] Deval Pandya [1]

Climate Change AI Workshop
ICLR 2024

[1] Vector Institute

VECTOR INSTITUTE | INSTITUT VECTEUR

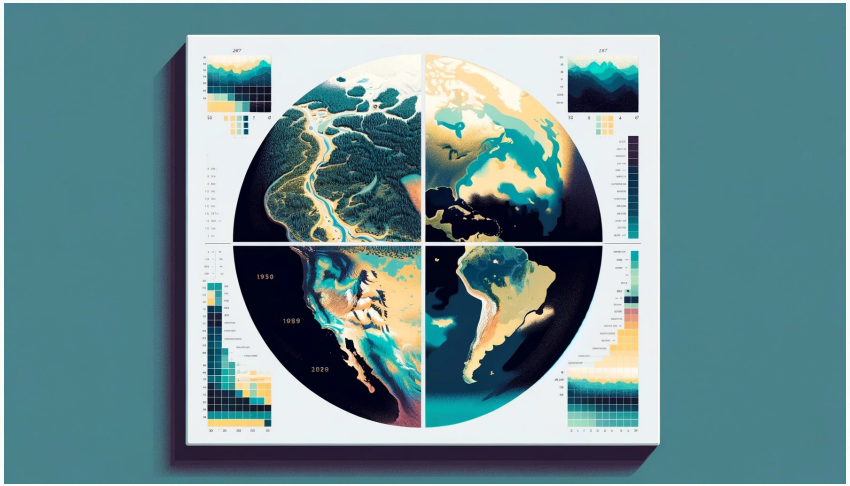
Background

Image segmentation is crucial in climate change research for analyzing satellite imagery.

- **Deforestation Monitoring**: Detect changes in forest cover over time, helping to pinpoint areas of illegal logging or deforestation.
- **Urban Planning**: Identify land use patterns, infrastructure, and green spaces, enabling better decision making and sustainable development.
- **Natural Disaster Response**: Quick analysis of areas affected by natural disasters to aid in efficient allocation of resources for relief efforts.
- And much more!

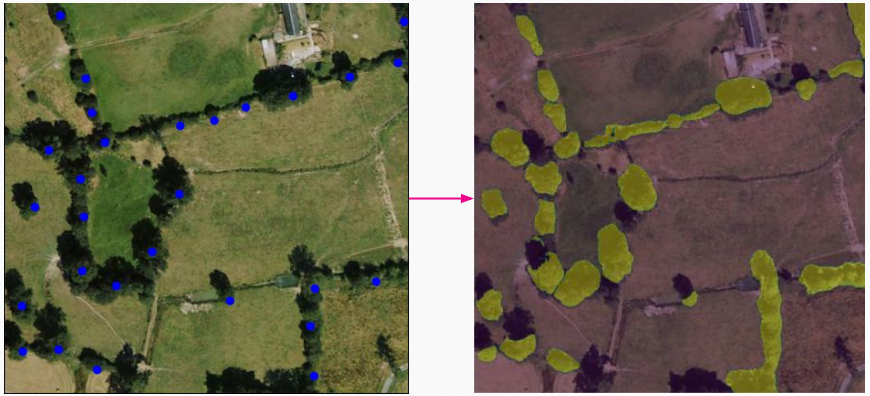
Background

- The advent of vision-based foundational models like the **Segment Anything Model (SAM)** opens new avenues in climate research and remote sensing (RS).
- SAM can perform zero-shot segmentation tasks from manually-crafted prompts



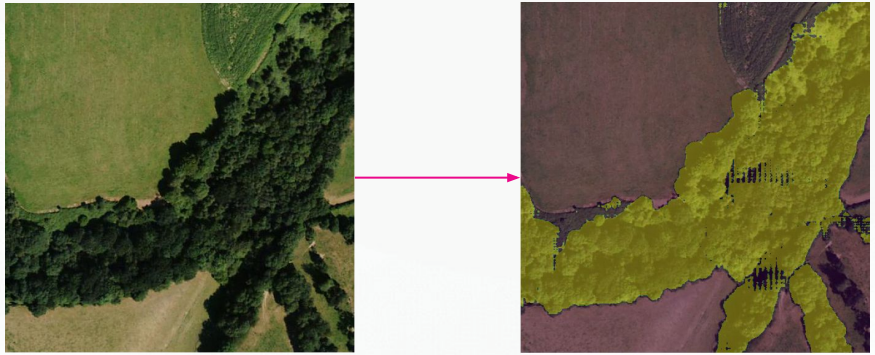
Motivation

- **The efficacy of SAM largely depends on the quality of input prompts.**
- This issue is particularly pronounced with RS data, which are inherently complex.
  - One would need to create **complex prompts** for each image, which typically involves selecting dozens of points or bounding boxes.
- **Hard to scale up!**

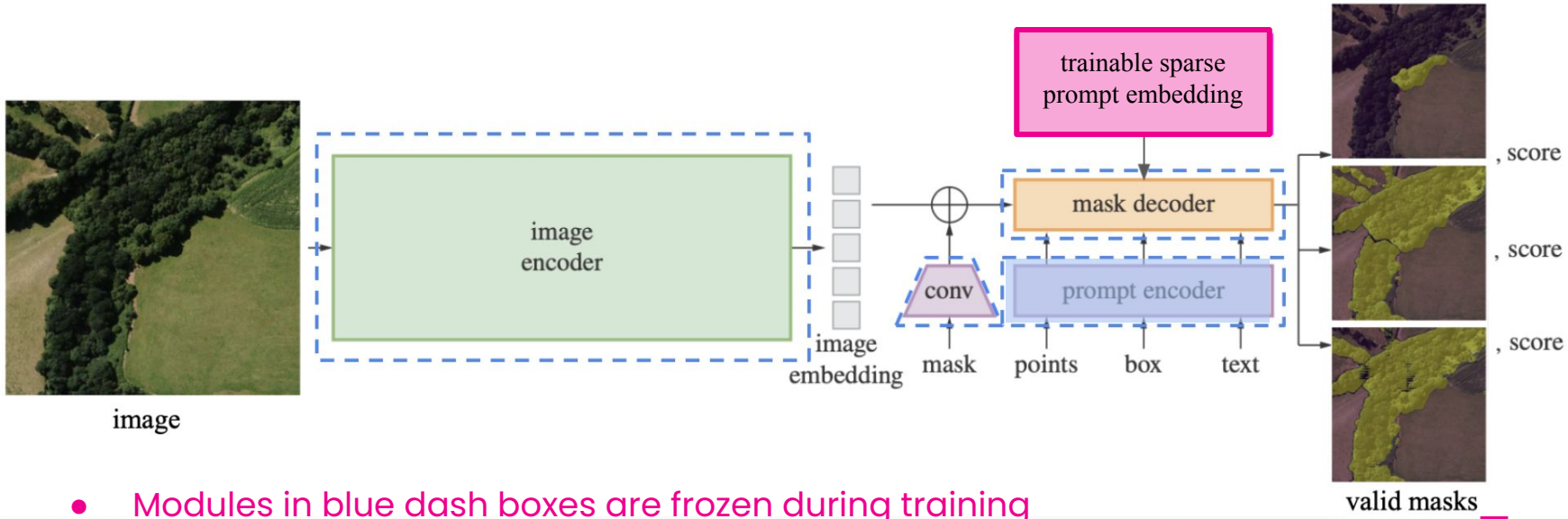


Prompt-Tuned SAM (PT-SAM)

- **A method that reduces the need for extensive manual input by incorporating a trainable, lightweight prompt embedding**
- **The embedding captures essential semantic information for specific objects**, enabling effective segmentation of unseen images without human intervention.
- Embedding training has **minimal dataset and hardware requirements** → Can be trained on CPU



PT-SAM Architecture



- Modules in blue dash boxes are frozen during training
- **Plug-and-play**: Users can easily replace the prompt encoder with various trained sparse prompt embedding

PT-SAM Training

1. Pre-compute image embeddings and generate fixed dense prompt embedding
2. Initialize the trainable sparse prompt embedding
3. For each image, pass the trainable sparse prompt embedding to the mask decoder along with the fixed dense prompt embedding as well as the image embedding to get generated masks
4. Compare the loss (BCE + DICE) between the generated masks and the corresponding ground truth masks
5. Perform linear regression on the trainable sparse prompt embedding based on the loss
6. Repeat from step 3 until convergence

PT-SAM Training

- **Minimal Hardware Requirement**: PT-SAM training experiments conducted on a single T4 GPU
  - If access to GPUs is limited, training can also be conducted on a CPU!
- **Minimal Dataset Requirement**:
  - Forest Prompt Embedding: Trained on 100 images with dimension of 500 x 500 pixels
  - Building Prompt Embedding: Trained on 1000 images with dimension of 1000 x 1000 pixels
- **Generalization**: Trained prompt embeddings can be applied to images of different sizes and different datasets.

Quantitative Results

- PT-SAM is baselined against SAM following the evaluation protocol used in the SAM paper
- The baseline method, referred to as Point-Prompted SAM (PP-SAM), uses the center point of the ground truth masks (or the closest point to the center on the masks) to prompt SAM.

Quantitative Results

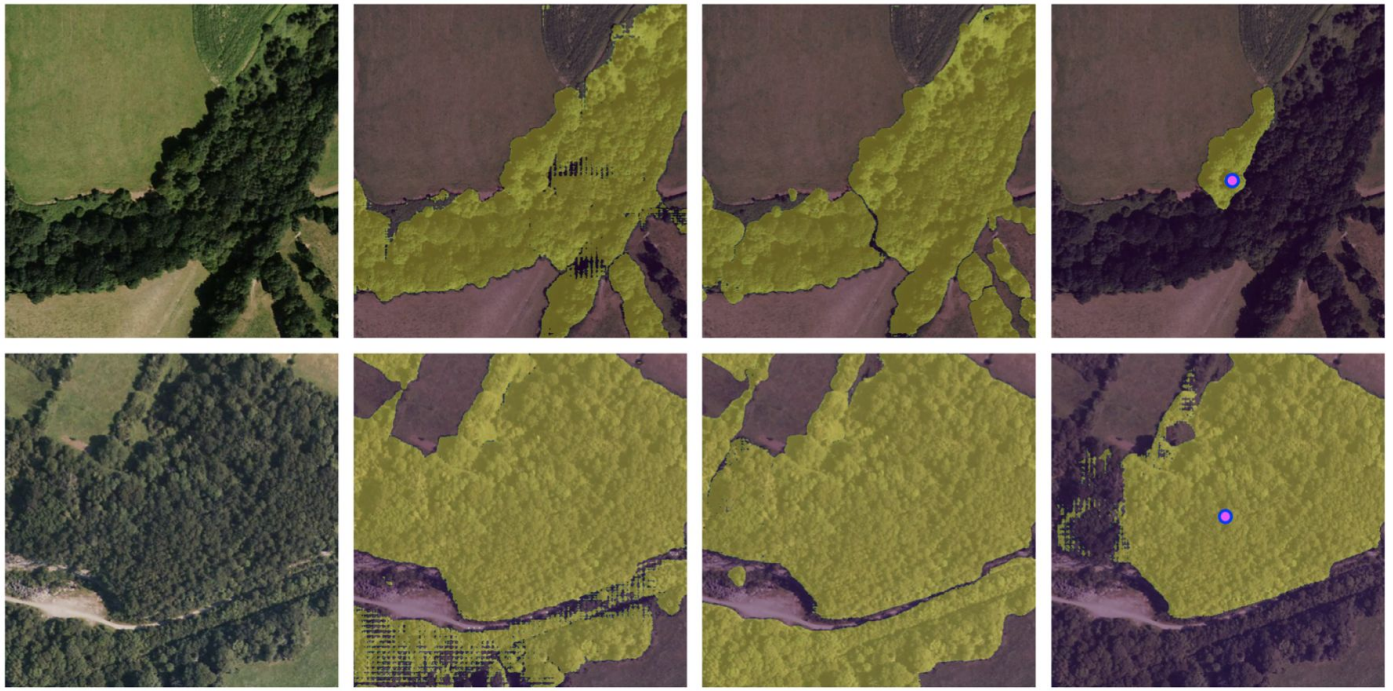
Table 1: Forest Embedding Evaluation Metrics

Average Scores	IoU	Accuracy	F1
PT-SAM (Ours)	**0.7639**	**0.8636**	**0.8558**
PP-SAM	0.4327	0.6220	0.4939

Table 2: Building Embedding Evaluation Metrics

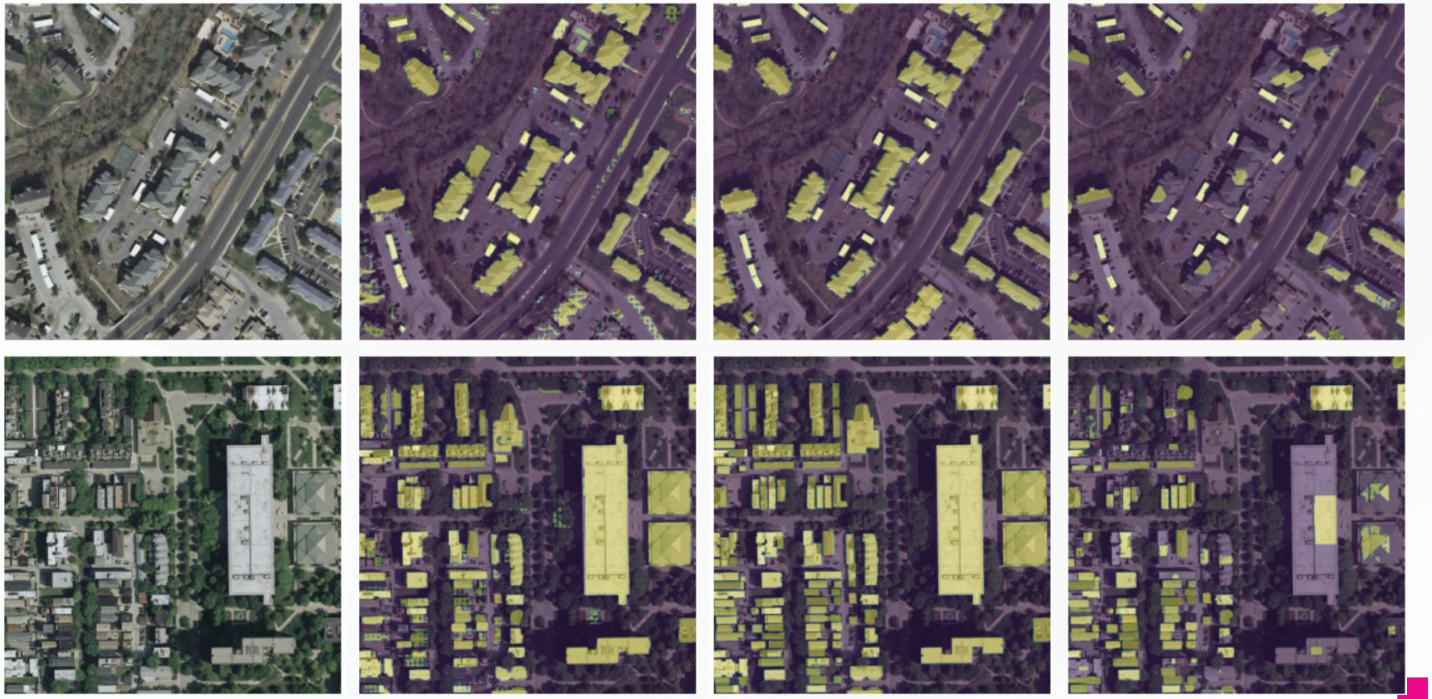
Average Scores	IoU	Accuracy	F1
PT-SAM (Ours)	**0.6686**	**0.9209**	**0.7990**
PP-SAM	0.3256	0.4399	0.4747

Qualitative Results - Forest



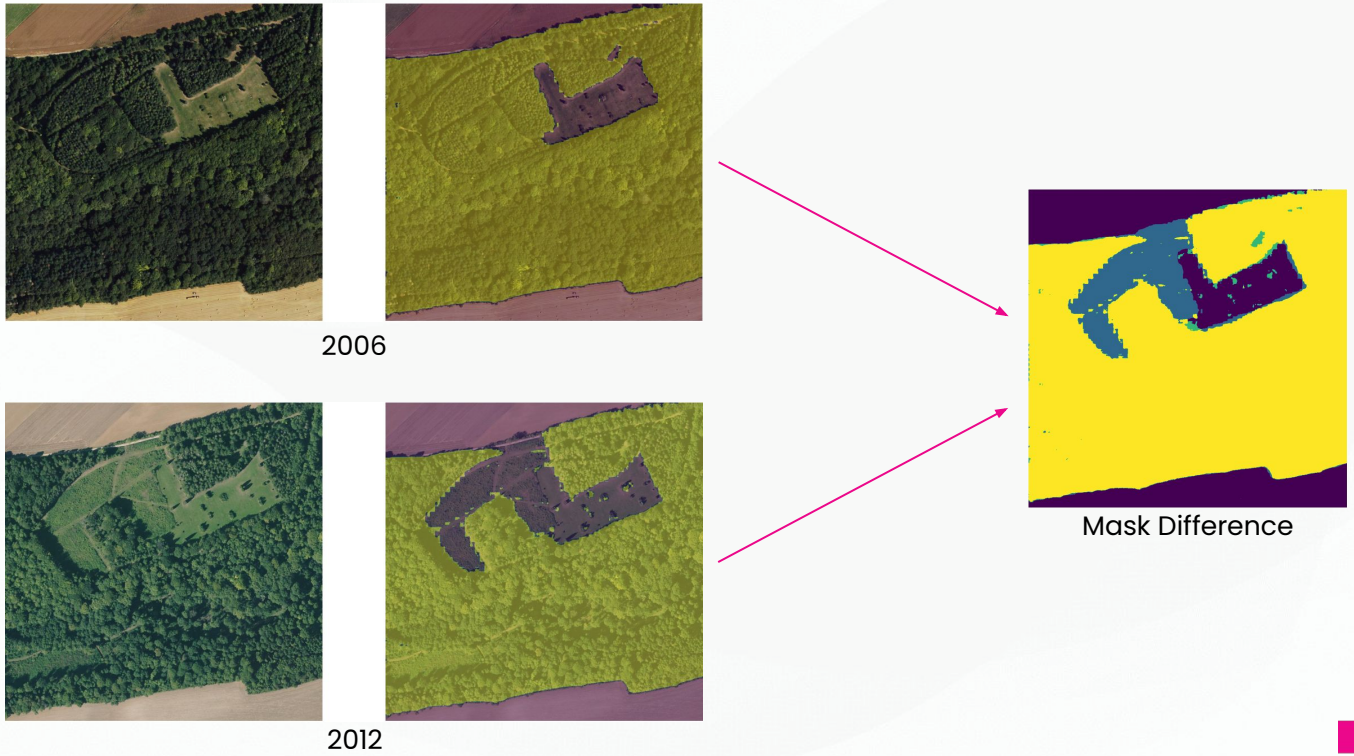
Original Image | PT-SAM | Ground Truth | PP-SAM

Qualitative Results - Buildings



Original Image | PT-SAM | Ground Truth | PP-SAM

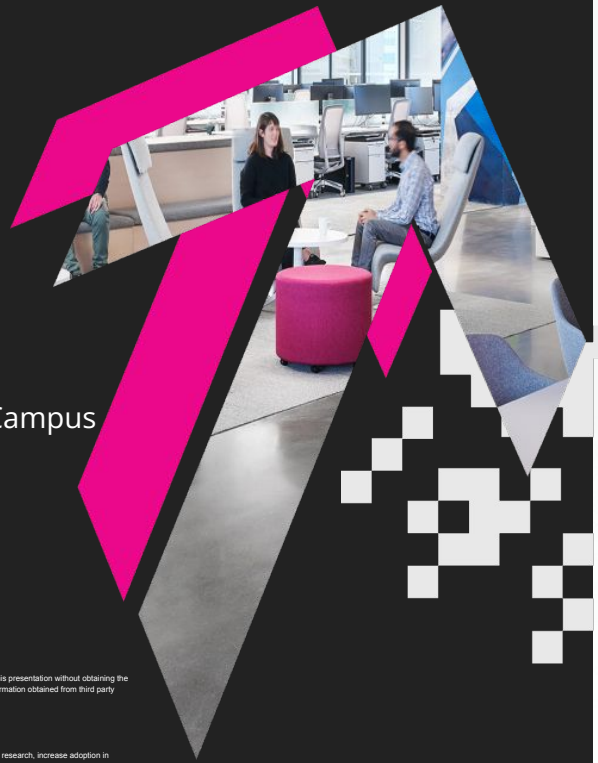
PT-SAM in Action: Deforestation Monitoring



Takeaways

- Introduced a **new lightweight approach** that combines the **zero-shot generalization** capabilities of SAM with supervised learning to **minimize manual input** in creating complex prompts for image segmentation.
- **Minimal hardware and dataset requirements** for training PT-SAM, making it accessible for widespread use.
- Demonstrated the applicability of PT-SAM for accurate and scalable image segmentation in remote sensing data, crucial for climate research.

Thank you!



vectorinstitute.ai
info@vectorinstitute.ai

Vector Institute
Schwartz Reisman Innovation Campus
W1140 -108 College Street,
Toronto, ON M5G 0C6
Canada

This presentation is for informational purposes only and Vector Institute is not, by means of this presentation, rendering professional advice or services or providing an opinion of any kind on any subject. No one should act upon or refrain from acting upon the information contained in this presentation without obtaining the advice of a qualified professional advisor. You are urged to contact a qualified professional advisor for guidance. Vector Institute does not warrant or guarantee the accuracy, currency, usefulness, or completeness of the information contained in this presentation, which may include information obtained from third party sources. Vector Institute shall not have any responsibility or owe any duty to any person in respect of this presentation or be responsible for any loss whatsoever sustained by any person who relies on this presentation.

All content in this presentation is either created by Vector Institute, or otherwise used with permission and is protected by copyright. All rights are reserved. No part of this presentation may be reproduced without the prior written permission of Vector Institute.

About Vector Institute
Launched in 2017, the Vector Institute works with industry, institutions, startups, and governments to build AI talent and drive research excellence in AI to develop and sustain AI-based innovation to foster economic growth and improve the lives of Canadians. Vector aims to advance AI research, increase adoption in industry and health through programs for talent, commercialization, and application, and lead Canada towards the responsible use of AI. Programs for industry, led by top AI practitioners, offer foundations for applications in products and processes, company-specific guidance, training for professionals, and connections to workforce-ready talent. Vector is funded by the Province of Ontario, the Government of Canada through the Pan-Canadian AI Strategy, and leading industry sponsors from across multiple sectors of Canadian Industry. For further information or media enquiries, please contact: media@vectorinstitute.ai

© The Vector Institute for Artificial Intelligence.