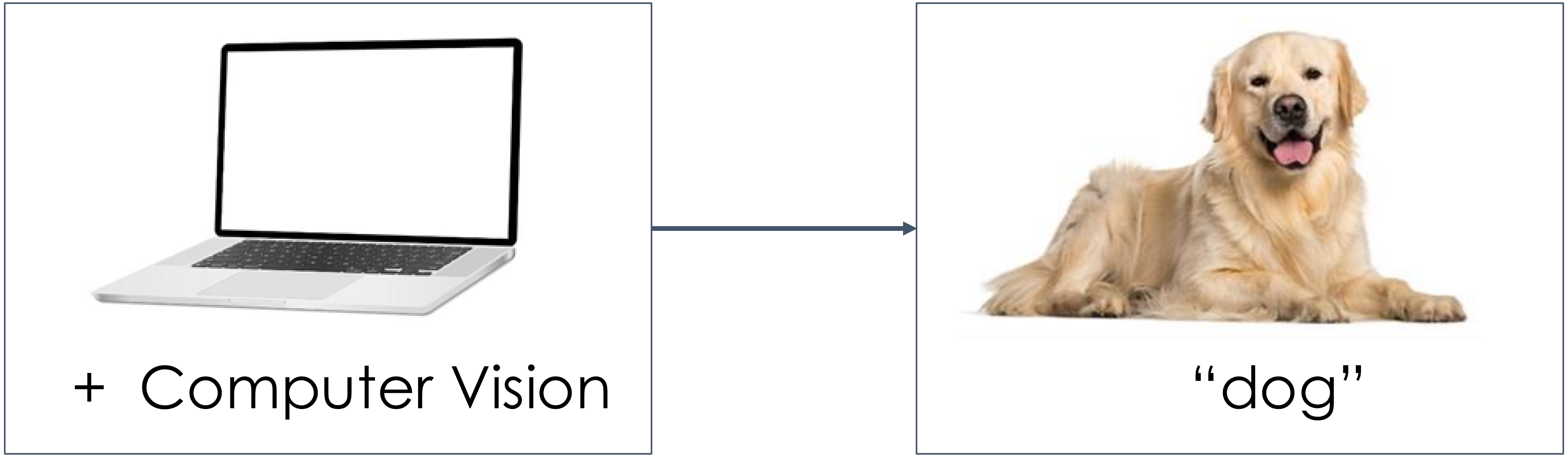
UMassAmherst | Manning College of Information & Computer Sciences

Computer Vision in the Wild

Rangel Daroya
University of Massachusetts Amherst

What is Computer Vision?

- Enables machines to **see the world** through visual inputs **similar to humans**
- Allows machines to make sense of images, videos, etc
  - Identifying objects
  - Classifying images
  - etc



What is Computer Vision?

- Enables machines to **see the world** through visual inputs **similar to humans**
- Allows machines to make sense of images, videos, etc
    - Identifying objects
    - Classifying images
    - etc

\+ Computer Vision

“tree”

Computer vision is slowly becoming a part of our daily lives

Apple's Face ID used in iPhone

Samsung's Circle to Search feature

AI Generated images (DeepFake)

University of Massachusetts Amherst

[1] https://www.differencebetween.net/technology/difference-between-facial-recognition-and-face-id/
[2] https://www.samsung.com/ph/support/mobile-devices/how-to-use-the-circle-to-search-feature-on-the-galaxy-s24/
[3] https://chameleonassociates.com/why-you-should-know-about-deepfake/

Machine Learning Overview

Step 1: Training

Data

Machine Learning Model

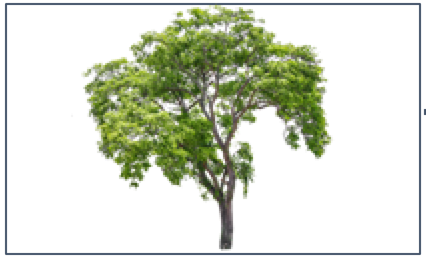
Label

"dog"

Computer vision uses **Machine Learning** to process and understand the visual world

Machine Learning Overview

Step 1: Training

Data		Label
	Machine Learning Model	"dog"
	Machine Learning Model	"cat"
	Machine Learning Model	"tree"

University of Massachusetts Amherst

Machine Learning Overview

Step 1: Training

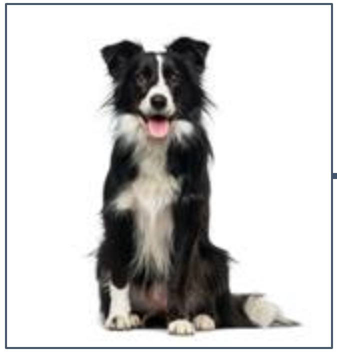
Machine Learning Model

"dog"

Data

Label

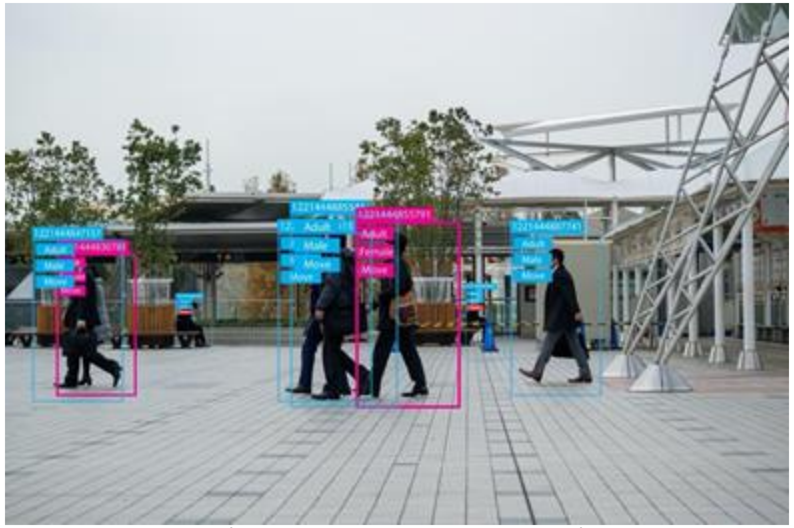
Step 2: Inference/Testing



Machine Learning Model

"dog"

What are common computer vision problems?

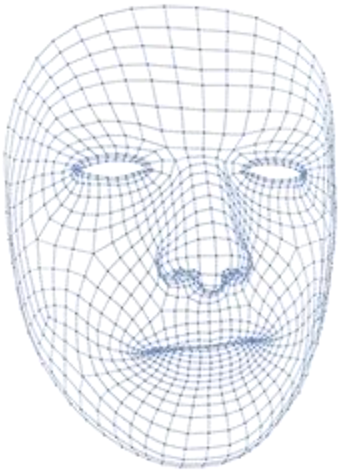


Object Detection

“dog”

“tree”

Image Classification



Face Recognition

Image Segmentation

Can we use computer vision to solve domain-specific problems?

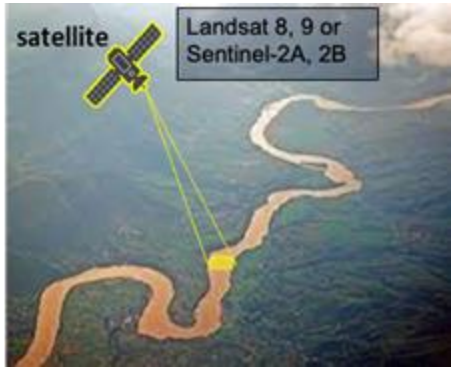
- **Domain-specific problems** are problems that refer to a specific area of expertise
  - For example:
    - water-related problems that hydrologists are concerned with
    - extra-terrestrial objects that physicists are interested in
- Why is this challenging?
  - Machine learning (ML) generally requires a lot of **labelled data**
  - Domain-specific problems generally don't have a lot of labelled data
  - **Labelled data are expensive** because they require inputs from experts

University of Massachusetts Amherst

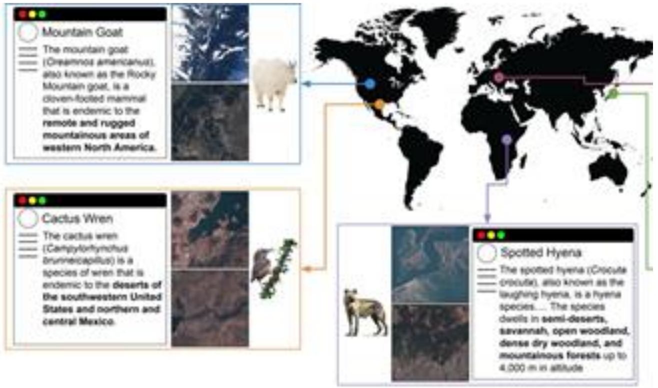
[1] https://engineeringonline.ucr.edu/blog/what-does-a-hydrologist-do/
[2] https://science.nasa.gov/mission/hubble/science/universe-uncovered/hubble-star-clusters/

Computer Vision in the Wild

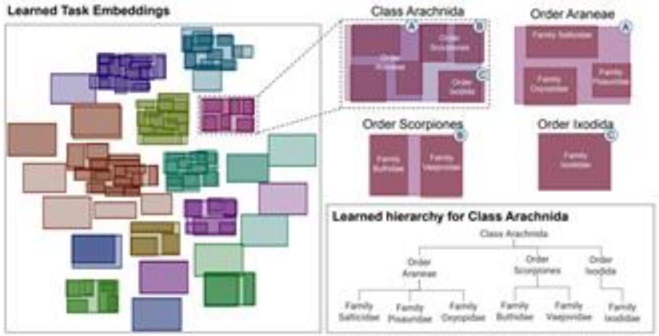
- In this talk, we will explore domain-specific applications of computer vision



1. Remote monitoring of global water quality using satellite images



2. Helping wildlife conservation efforts through wildlife habitat information and satellite imagery

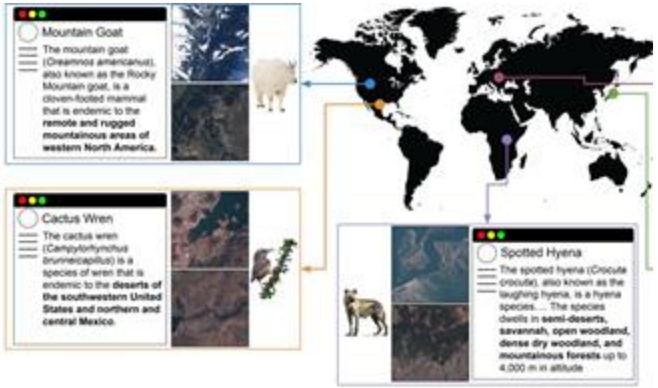


3. Modeling relationships between domain-specific tasks for efficient ML model training

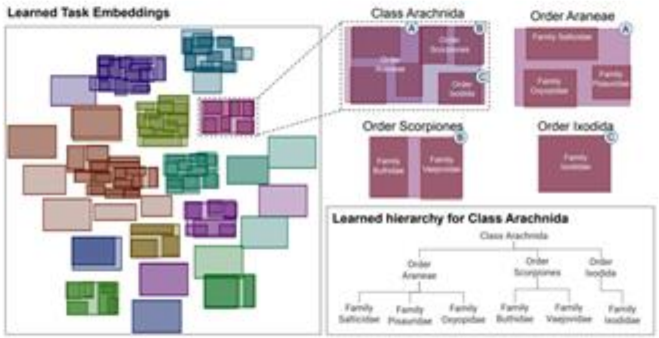
Computer Vision in the Wild

- In this talk, we will explore domain-specific applications of computer vision

1. Remote monitoring of global water quality using satellite images



2. Helping wildlife conservation efforts through wildlife habitat information and satellite imagery



3. Modeling relationships between domain-specific tasks for efficient ML model training

UMassAmherst | Manning College of Information & Computer Sciences

A hosted analytic collaborative framework for global river water quantity and quality from SWOT, Landsat, and Sentinel-2

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji

University of Pittsburgh

University of Massachusetts Amherst

The University of North Carolina at Chapel Hill

Overview

A hosted analytic collaborative framework for global river water quantity and quality from SWOT, Landsat, and Sentinel-2

University of Massachusetts Amherst

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

A hosted analytic collaborative framework for global river water quantity and quality from **SWOT, Landsat, and Sentinel-2**

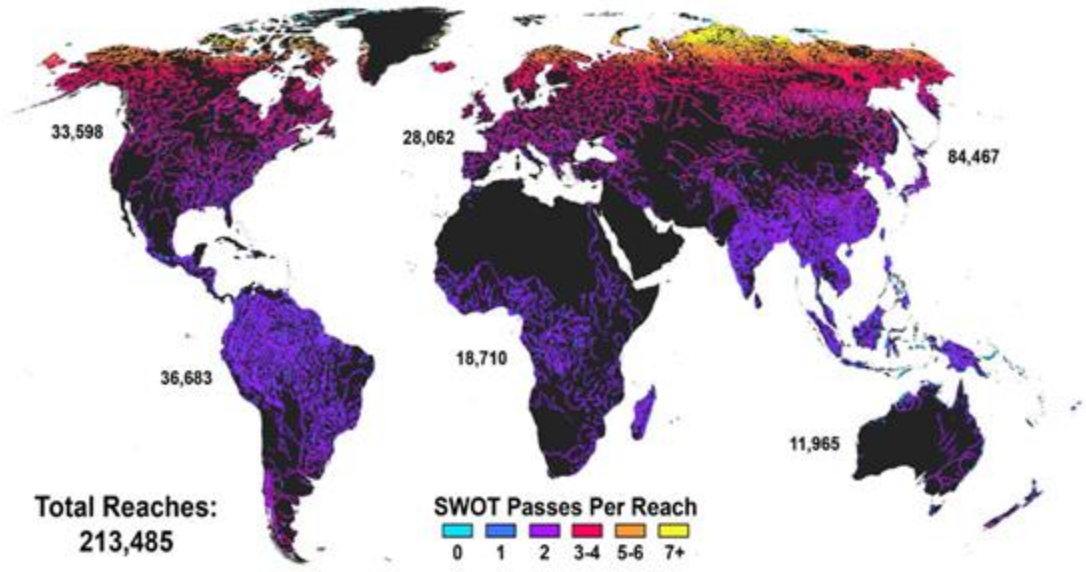
- SWOT, Landsat, and Sentinel are **NASA satellite missions** with global coverage
- SWOT aims to survey Earth's waters to observe **fine details** of oceans and measure changes in terrestrial waters
- Landsat and Sentinel are used to acquire satellite imagery over Earth (lower resolution than SWOT) for **general remote sensing applications**

University of Massachusetts Amherst

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

Overview

A hosted analytic collaborative framework for **global river water quantity and quality** from SWOT, Landsat, and Sentinel-2

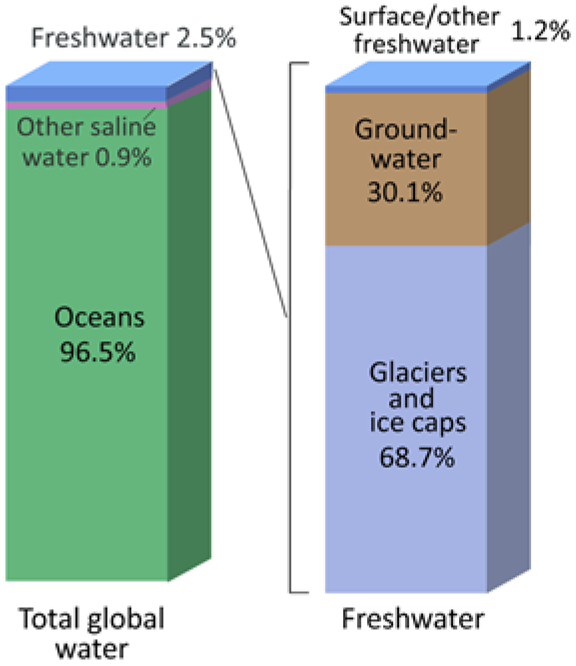


- **River water quantity:** SWOT can be used to derive discharge (amount of water flowing in a river)
- **River water quality:** To measure water quality we use the amount of **suspended sediment concentration (SSC)**

University of Massachusetts Amherst

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

Why rivers and sediments?

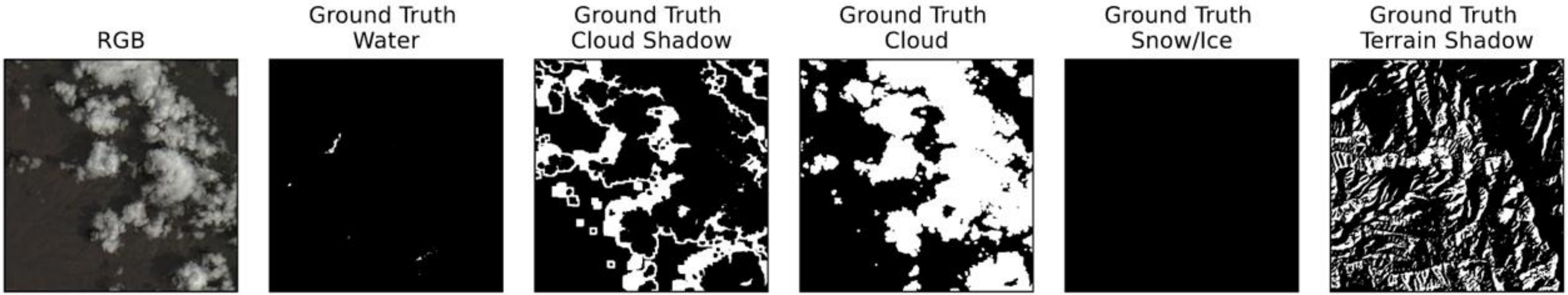


- Rivers are integral to communities as a source of freshwater
  - Drinking water
  - Irrigation
  - Power (through dams)
  - Homes for fish, wildlife, and plants
- Sediment is an important measure of rivers that is observable from space
  - Product of erosion
  - Needed for coastal resilience (e.g., for flooding)
  - Impacts hydropower efficiency

University of Massachusetts Amherst

[1] https://www.usgs.gov/media/images/distribution-water-and-above-earth

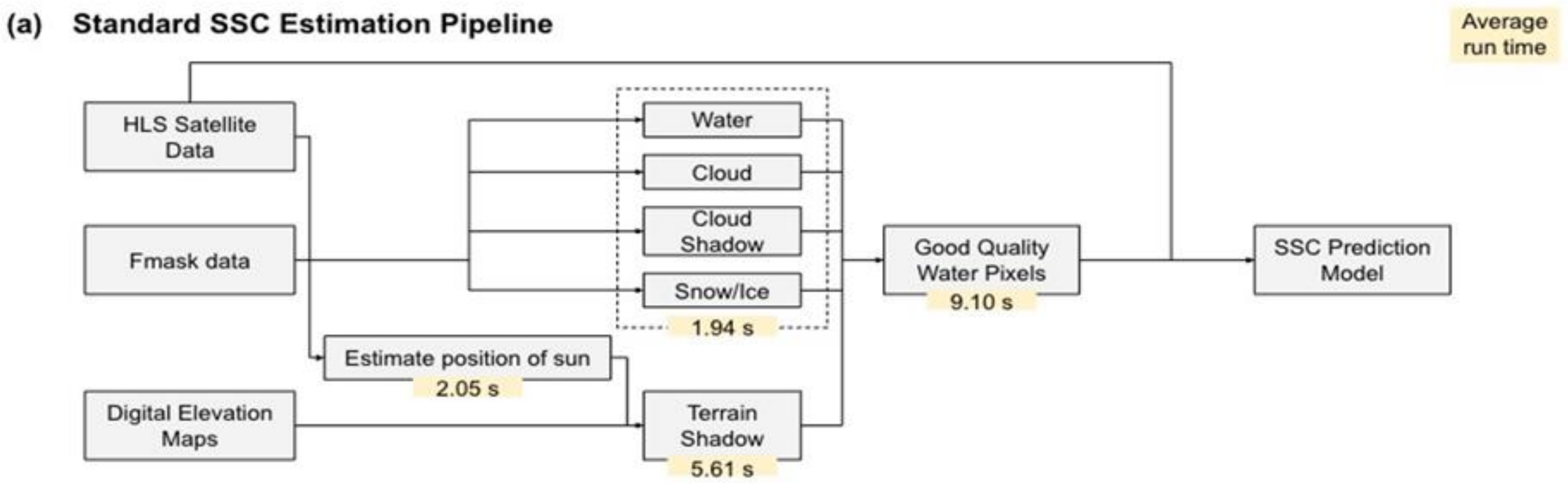
Estimating suspended sediments require finding water from satellite images



- Finding water is an important prerequisite for sediment estimation
- We want to find **good quality water pixels** to have an accurate model for SSC predictions
- Good quality water pixels are **water pixels** not covered by
  - Cloud shadows
  - Clouds
  - Ice/snow
  - Terrain shadows

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

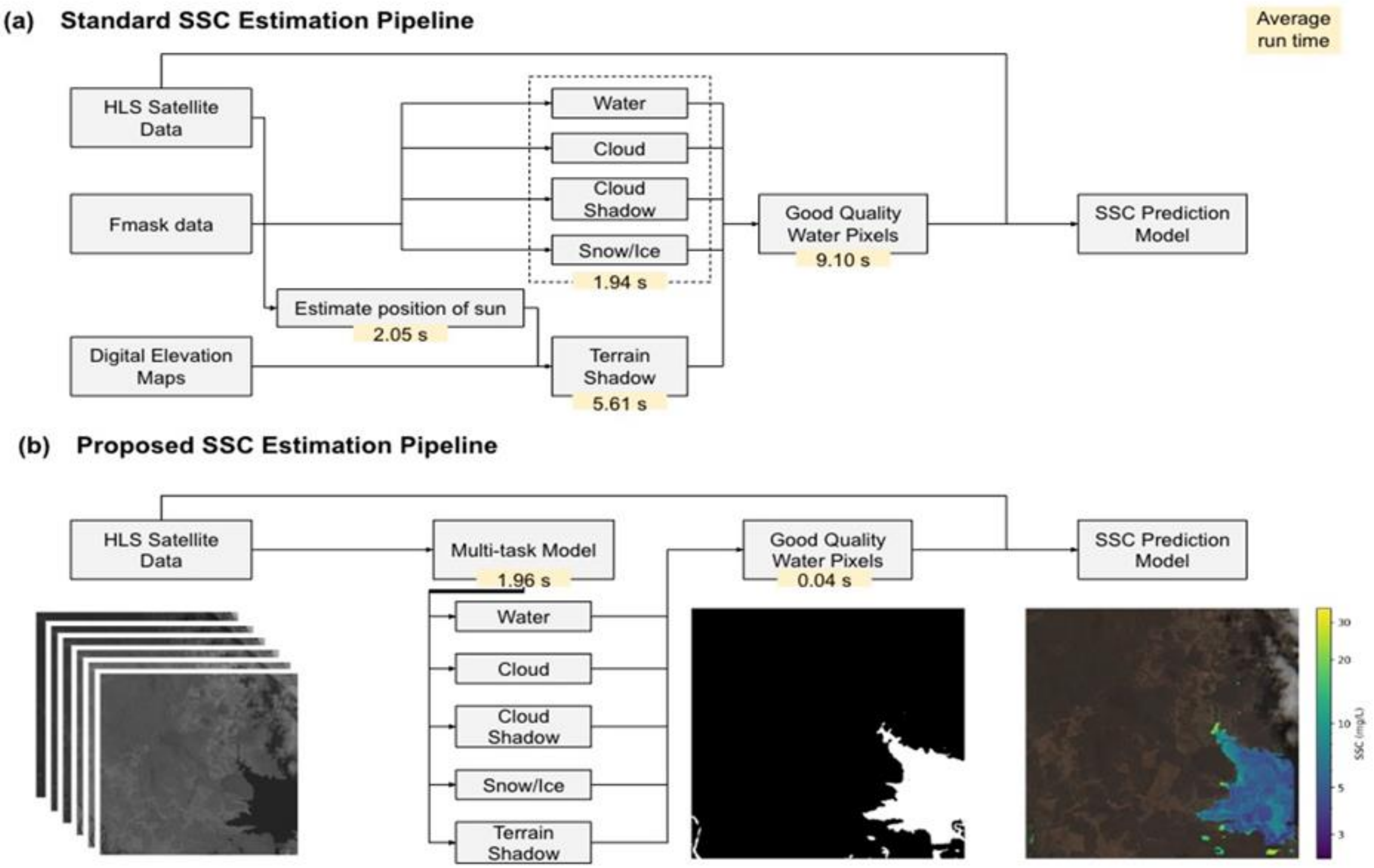
Standard pipelines are resource-intensive



University of Massachusetts Amherst

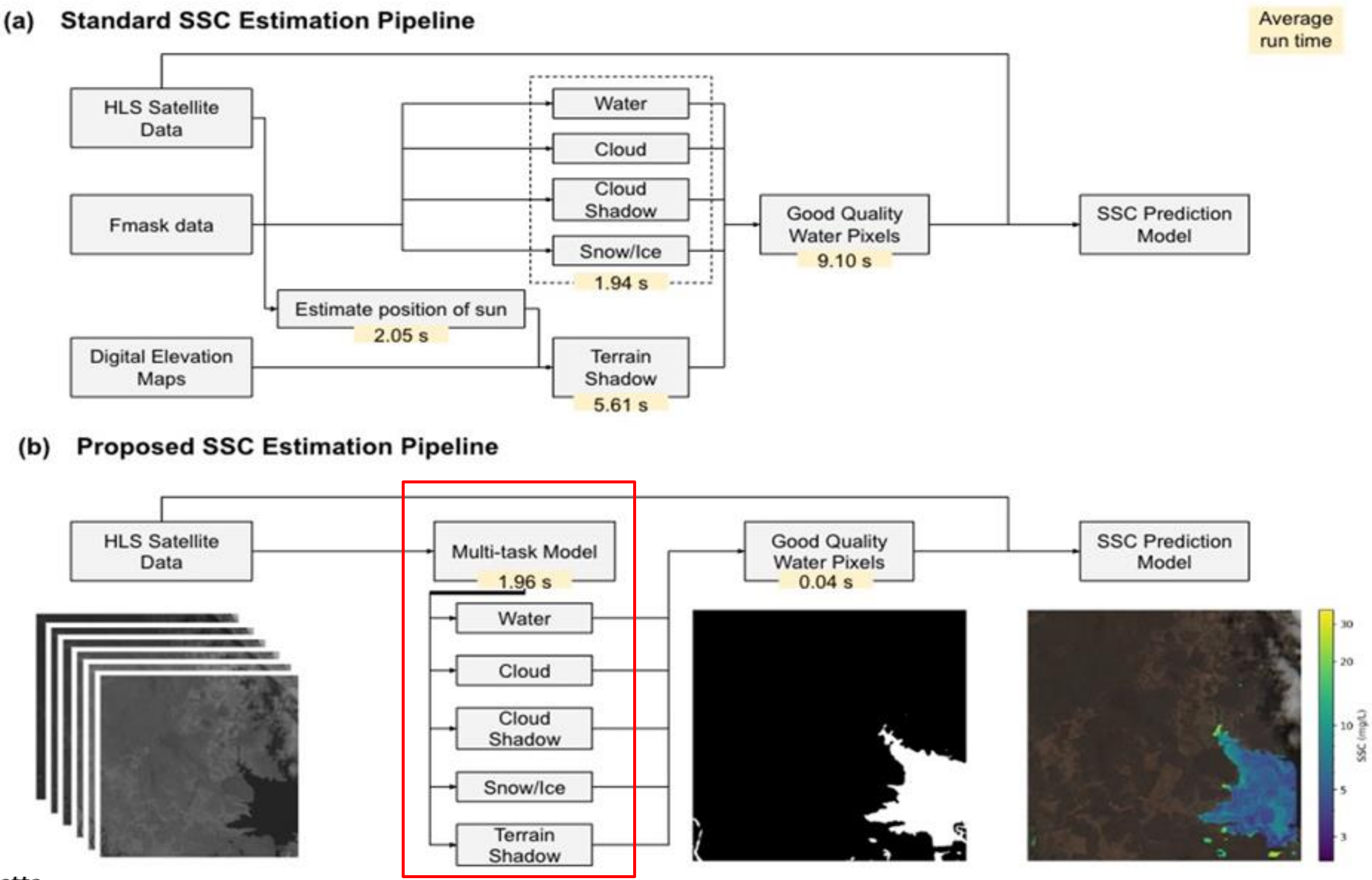
Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

We propose a multi-task model for a more accurate and efficient SSC pipeline



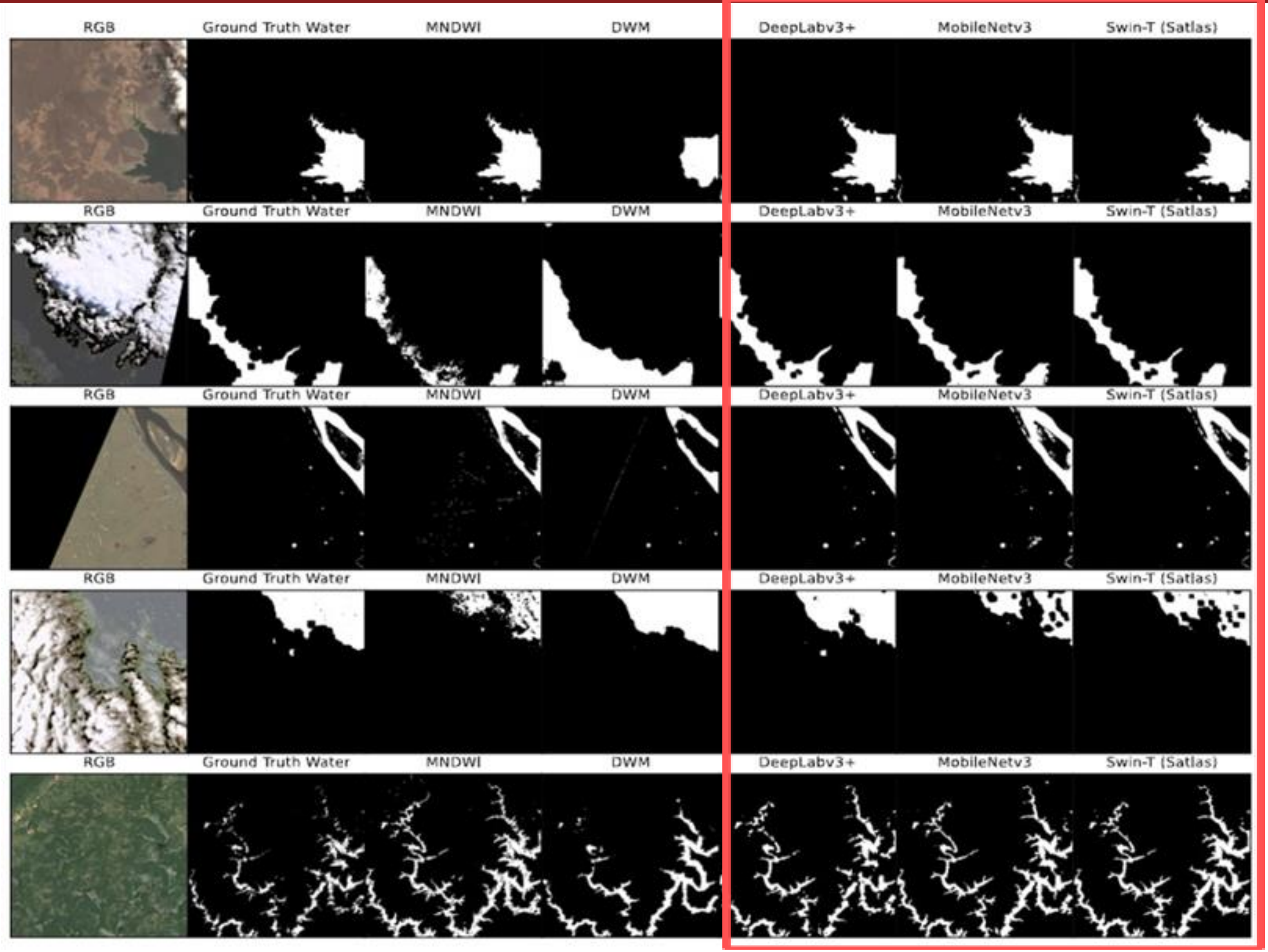
Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

We propose a multi-task model for a more accurate and efficient SSC pipeline



Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

We validate our proposed model on different architectures and pre-training methods



Proposed multi-task models have qualitatively better results and quantitatively better results

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

University of Massachusetts Amherst

We validate on manually labeled cloud data from USGS personnel

- We show that we **outperform** previous methods **by at least 12% and 6% F1 Score** for cloud shadow and clear predictions, respectively

	Model	Cloud (↑)	Cloud Shadow (↑)	Clear (↑)
Baselines	LANA (Zhang et al., 2024)	92.42%	57.53%	89.02%
	Fmask (Qiu et al., 2019)	89.81%	45.42%	88.09%
	U-Net Wieland (Wieland et al., 2019)	87.68%	52.06%	86.19%
Multi-task Models	DeepLabv3+ (ImageNet pre-trained)	92.64%	<u>65.79%</u>	95.54%
	MobileNetv3 (ImageNet pre-trained)	**93.70%**	63.60%	<u>95.77%</u>
	SegNet (ImageNet pre-trained)	91.19%	57.64%	95.19%
	ResNet50 (Satlas pre-trained)	85.78%	63.67%	92.77%
	Swin-T (Satlas pre-trained)	<u>92.96%</u>	**69.56%**	**95.80%**
	Swin-T (ImageNet pre-trained)	82.73%	4.32%	92.49%
	Vit-B/16 (ImageNet pre-trained)	59.89%	0.01%	88.16%
	Vit-B/16 (Prithvi pre-trained)	81.38%	6.94%	91.52%

Table. Performance of cloud masking on manually labeled LANA dataset. Overall performance is measured with F1 Score.

University of Massachusetts Amherst

[1] Zhang, Hankui K., Dong Luo, and David P. Roy. "Improved Landsat Operational Land Imager (OLI) Cloud and Shadow Detection with the Learning Attention Network Algorithm (LANA)." Remote Sensing 16.8 (2024): 1321.

We validate our proposed model on different architectures and pre-training methods

- We outperform baseline methods by almost **9% F1 Score** on water masking

Method	Pre-training	Model Type	F1 Score (↑)	Precision (↑)	Recall (↑)	IoU (↑)
MNDWI			58.43%	78.92%	46.39%	41.28%
DWM		CNN	82.21%	78.54%	86.24%	69.79%
DeepLabv3+	ImageNet	CNN	89.67%	87.91%	91.50%	81.27%
MobileNetv3	ImageNet	CNN	88.18%	85.16%	91.42%	78.86%
SegNet	ImageNet	CNN	83.47%	82.94%	84.01%	71.63%
ResNet50	Satlas	CNN	81.33%	78.76%	84.08%	68.54%
Swin-T	Satlas	Transformer	**91.10%**	**90.62%**	**91.58%**	**83.65%**
Swin-T	ImageNet	Transformer	80.73%	77.88%	83.80%	67.69%
ViT-B/16	ImageNet	Transformer	82.56%	81.15%	84.03%	70.30%
ViT-B/16	Prithvi	Transformer	76.61%	74.60%	78.74%	62.09%

University of Massachusetts Amherst

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

We validate our proposed model on different architectures and pre-training methods

- Runtime is **30x faster than baseline**, making it possible to run the model daily for frequent monitoring of rivers

	Runtime on 1 sample (s)	Runtime on 400k samples (days)	Improvement (%)
Standard SSC Pipeline	18.757	86.84	-
DeepLabv3+ (ImageNet pre-trained)	2.002	9.27	89.33%
MobileNetv3 (ImageNet pre-trained)	**0.601**	**2.78**	**96.80%**
SegNet (ImageNet pre-trained)	2.259	10.46	87.96%
ResNet50 (Satlas pre-trained)	7.209	33.38	61.57%
SwinT (Satlas pre-trained)	6.260	28.98	66.62%
SwinT (ImageNet pre-trained)	<u>1.254</u>	<u>5.80</u>	<u>93.32%</u>
ViT-B/16 (ImageNet pre-trained)	2.450	11.34	86.94%
ViT-B/16 (Prithvi pre-trained)	3.493	16.17	81.38%

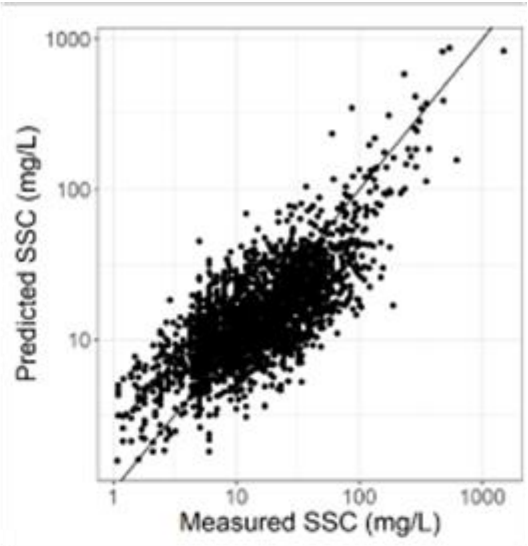
University of Massachusetts Amherst

Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

Conclusion

A hosted analytic collaborative framework for global river water quantity and quality from SWOT, Landsat, and Sentinel-2

- We can observe SSC over any water in satellite imagery
  - Introduces performance and speed improvements
  - Possible to run at a high frequency over time and space



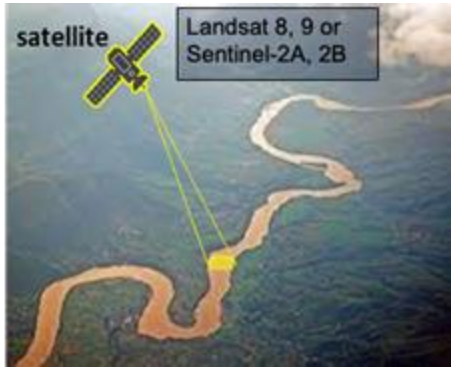
arxiv.org/abs/2412.08545

University of Massachusetts Amherst

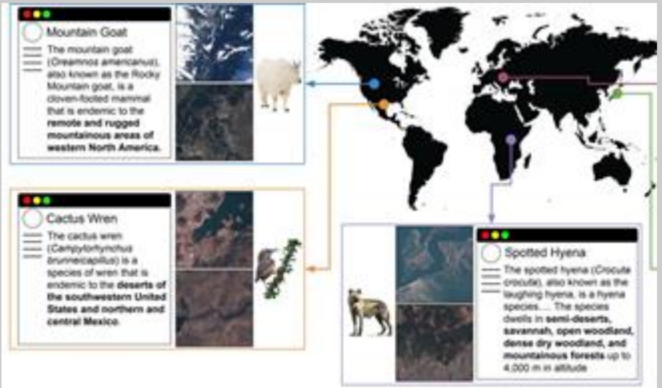
Rangel Daroya, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.

Computer Vision in the Wild

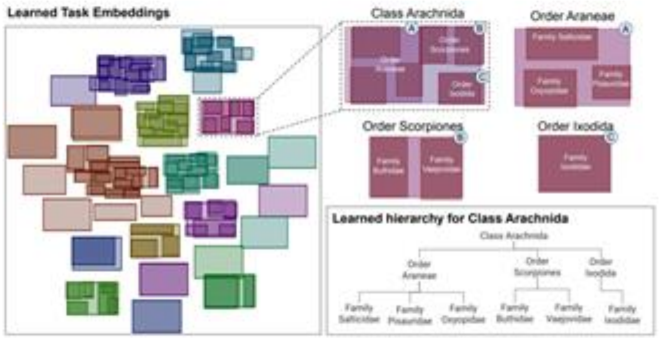
- In this talk, we will explore domain-specific applications of computer vision



1. Remote monitoring of global water quality using satellite images



2. Helping wildlife conservation efforts through wildlife habitat information and satellite imagery



3. Modeling relationships between domain-specific tasks for efficient ML model training

UMassAmherst | Manning College of Information & Computer Sciences

WildSAT: Learning Satellite Image Representations from Wildlife Observations

Rangel Daroya[1], Elijah Cole[2], Oisin Mac Aodha[3], Grant Van Horn[1], Subhransu Maji[1]

Recall: Machine Learning Overview

Step 1: Training

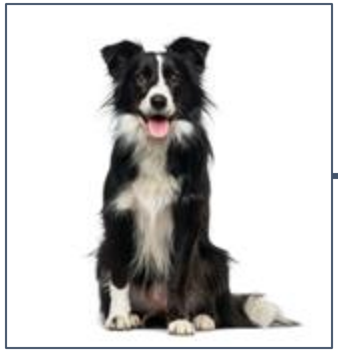
Machine Learning Model

"dog"

Data

Label

Step 2: Inference/Testing

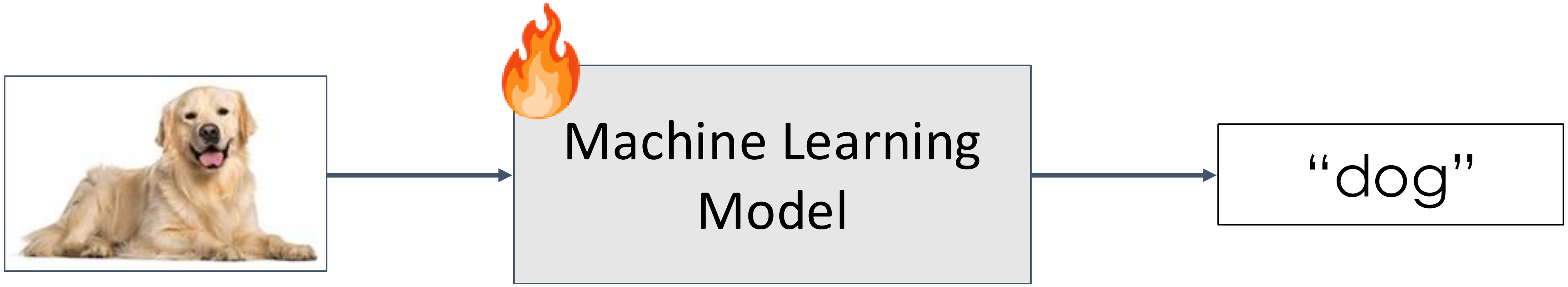


Machine Learning Model

"dog"

University of Massachusetts Amherst

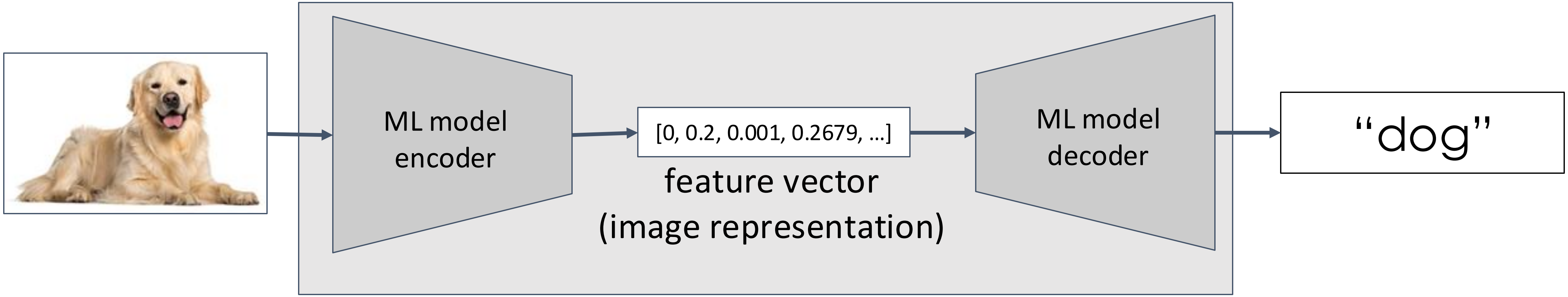
Model Feature Vectors



Data

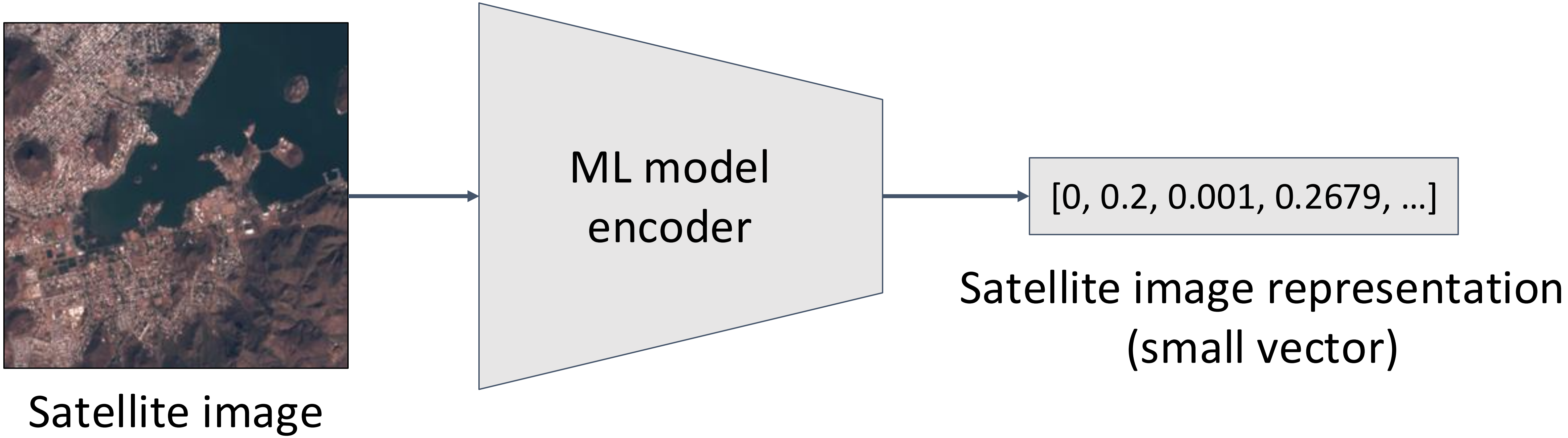
Label

The input to an ML model can be compressed to a **vector**



Satellite image representations

Satellite image representations refer to the encoded **satellite image** using a given ML model



Goal: train <u>any</u> ML model to be better at producing satellite image representations (better means improved performance on satellite image tasks)

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

Overview

- **Main question:** can we use the distribution of wildlife to improve satellite image representation?
- To do this, we need a dataset with the following:
  - **Location:** Latitude, longitude
  - **Satellite images** at the given location
  - **Text** that describe the **species** present at the location and corresponding information about their habitat

University of Massachusetts Amherst

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

Wildlife observation data can be used to learn about the different habitats

Data from iNaturliast [1], combined with Wikipedia [2] and Sentinel [3] data can be used

Why use wildlife observation data?

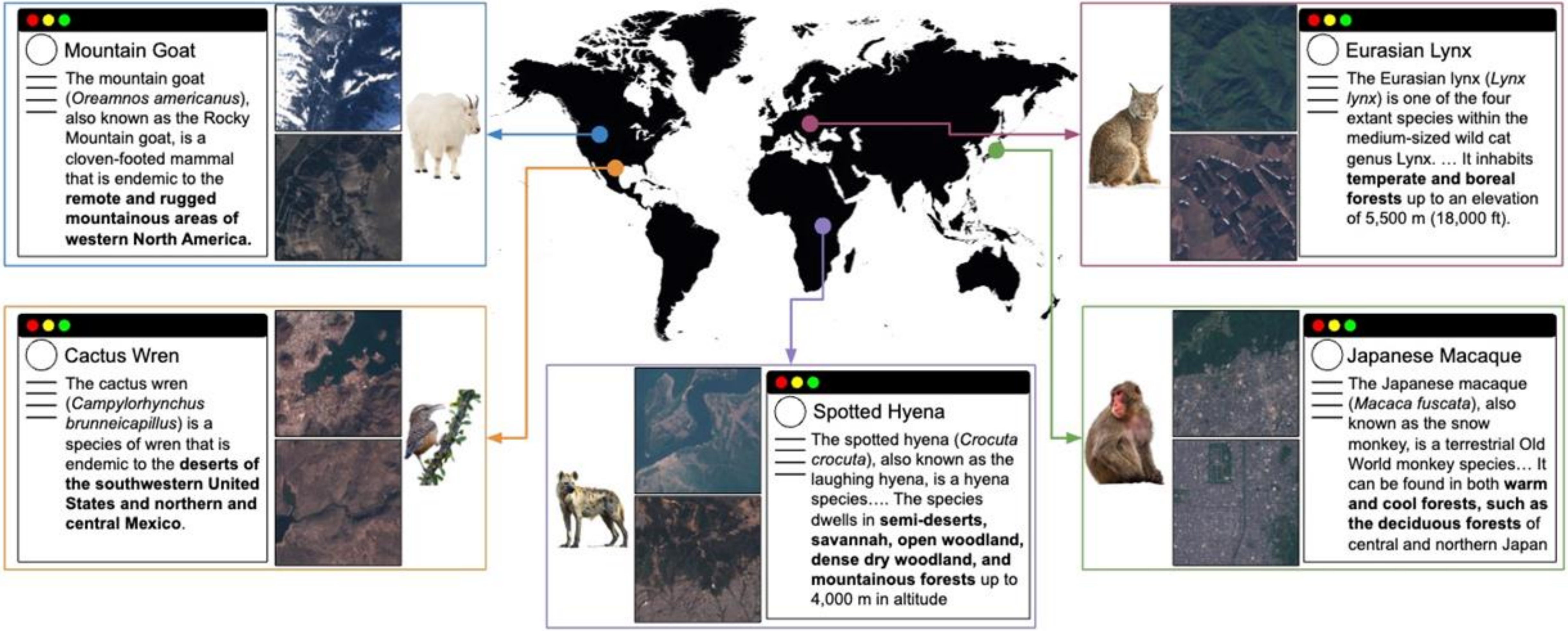
- Associated text of the wildlife contains information about **the habitat and the types of environment** each species prefer
- These habitat descriptions can then supplement and further describe satellite imagery

[1] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778, 2018
[2] Wikipedia. https://www.wikipedia.org. Accessed on 2024-11-14.
[3] ESA. Sentinel-1-missions-sentinel online-sentinel online. Eur. Sp. Agency, 2022.

University of Massachusetts Amherst

Wildlife observation data can be used to learn about the different habitats

Data from iNaturliast [1], combined with Wikipedia [2] and Sentinel [3] data can be used



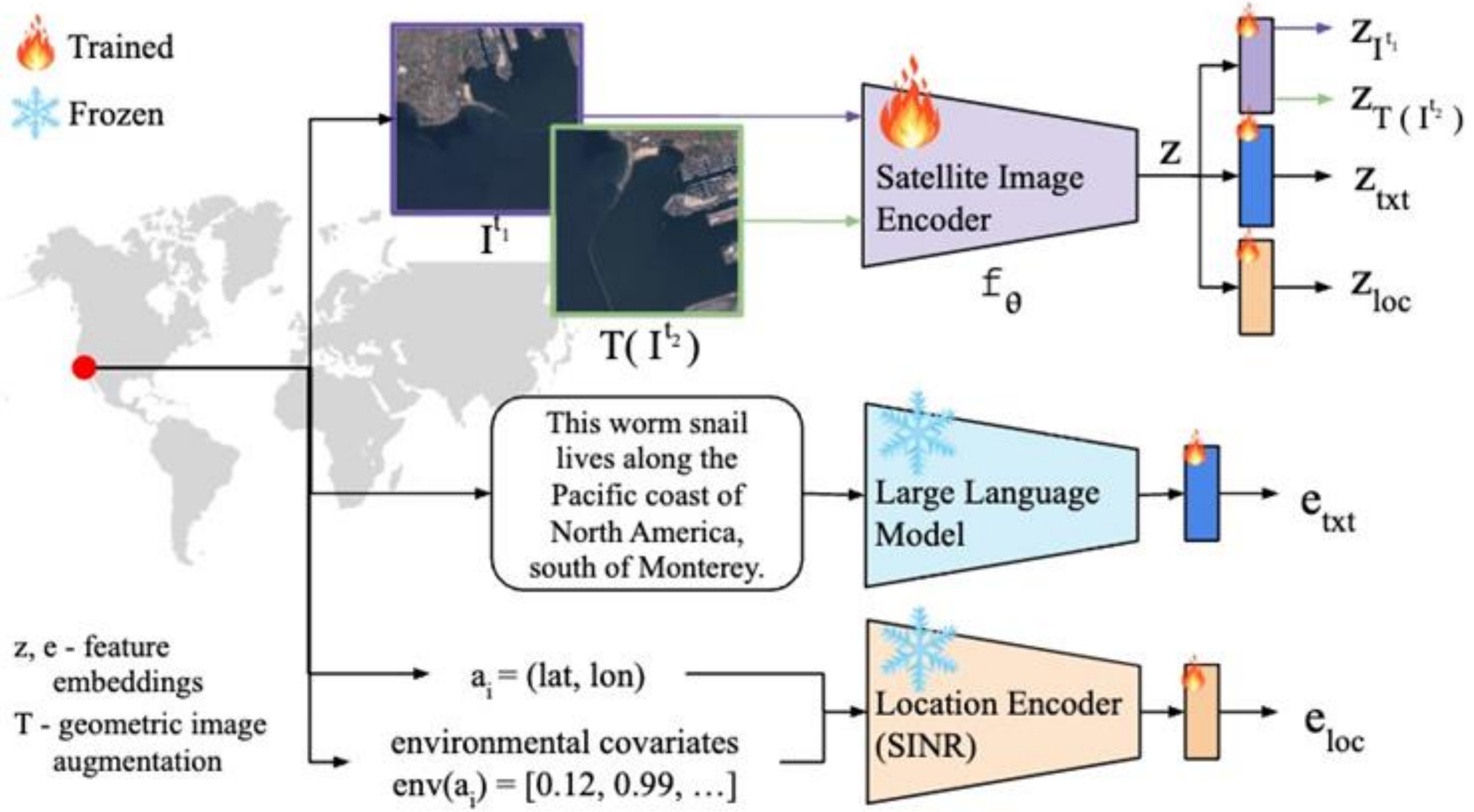
University of Massachusetts Amherst

[1] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778, 2018

[2] Wikipedia. https://www.wikipedia.org. Accessed on 2024-11-14.

[3] ESA. Sentinel-1-missions-sentinel online-sentinel online. Eur. Sp. Agency, 2022.

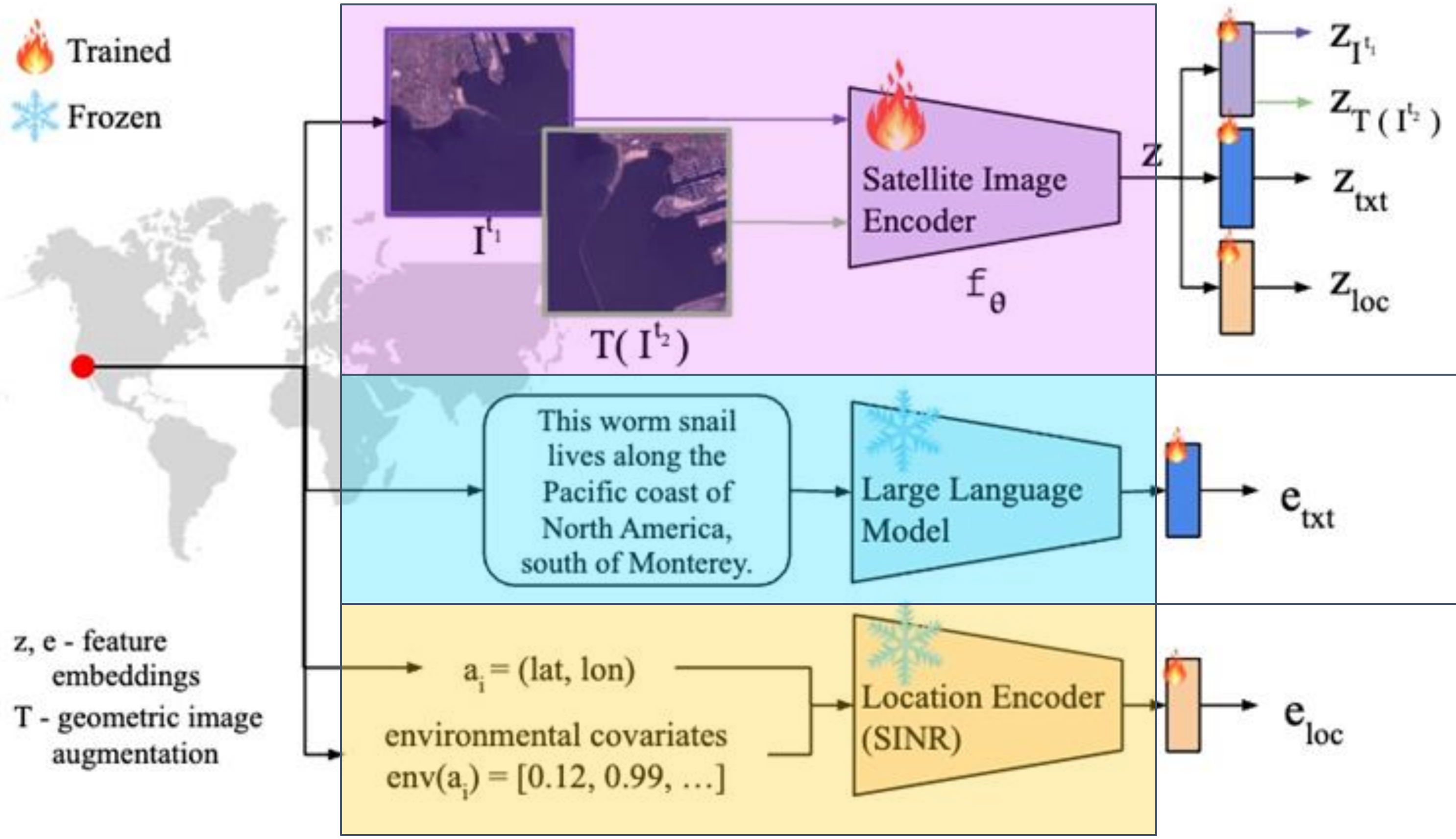
WildSAT Architecture (Contrastive Training)



University of Massachusetts Amherst

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT Architecture (Contrastive Training)



Image

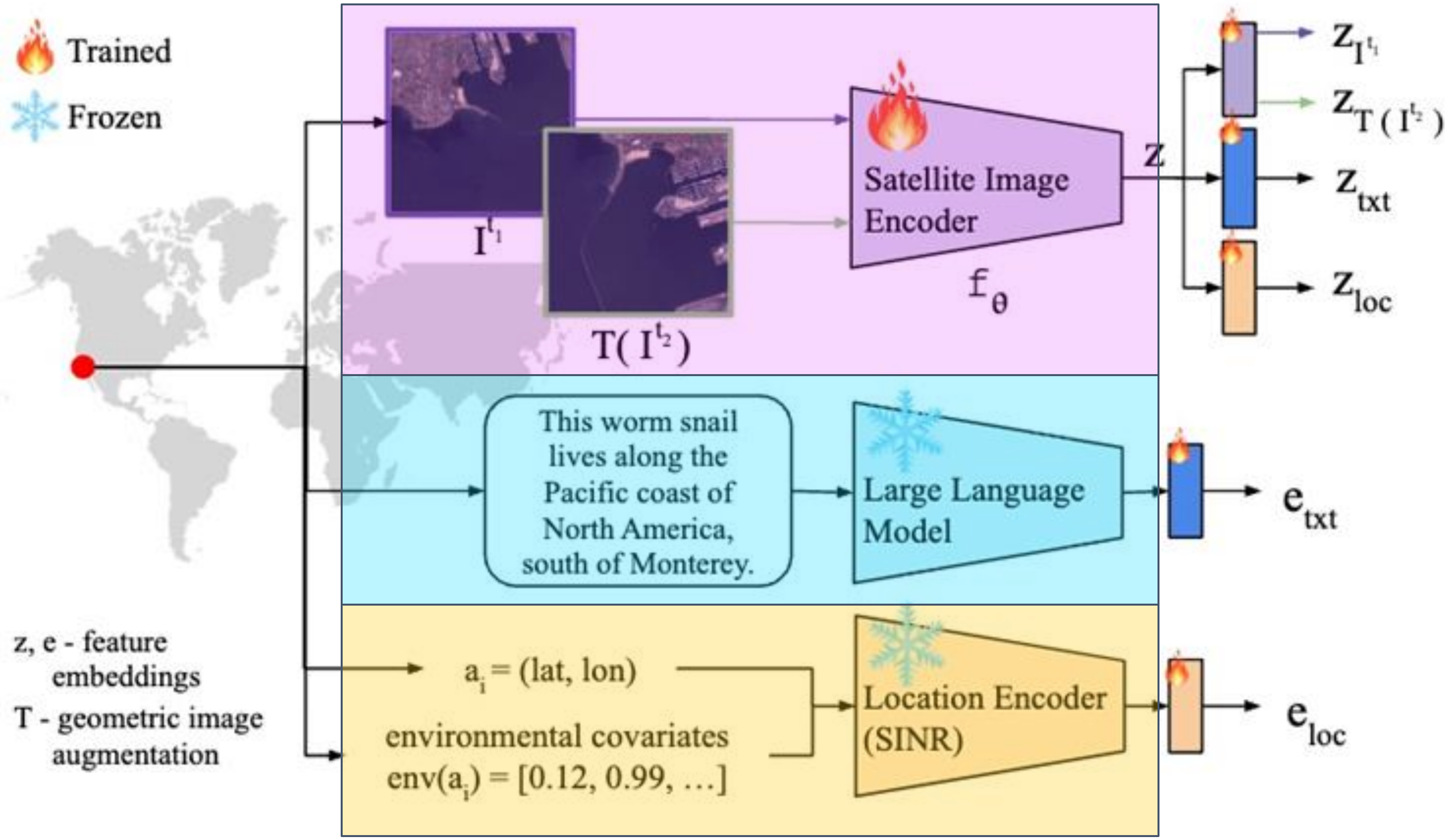
Text

Location

University of Massachusetts Amherst

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT Architecture (Contrastive Training)



Training involves 3 objectives that cover the **3 modalities**:

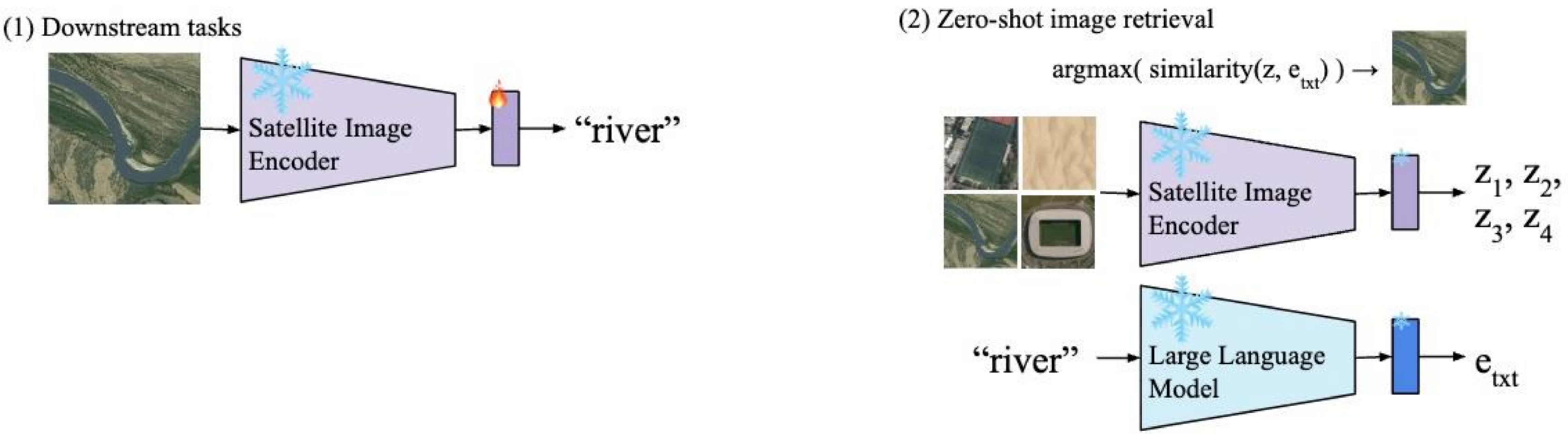
1. Image
2. Text
3. Location

Contrastive loss is used for each of the objectives

- encourages similar data to have similar representations

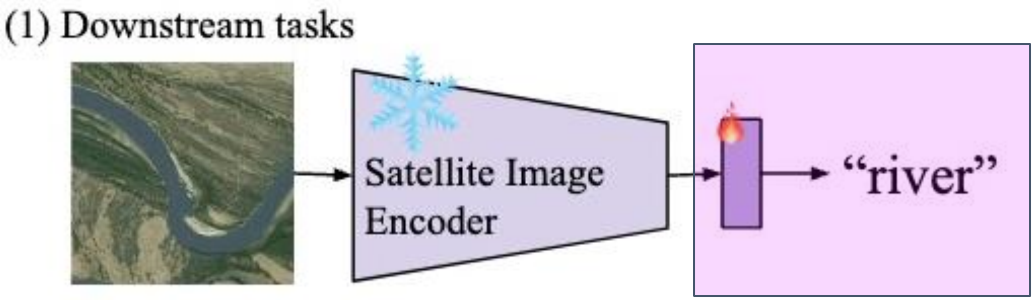
Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT at Inference Time



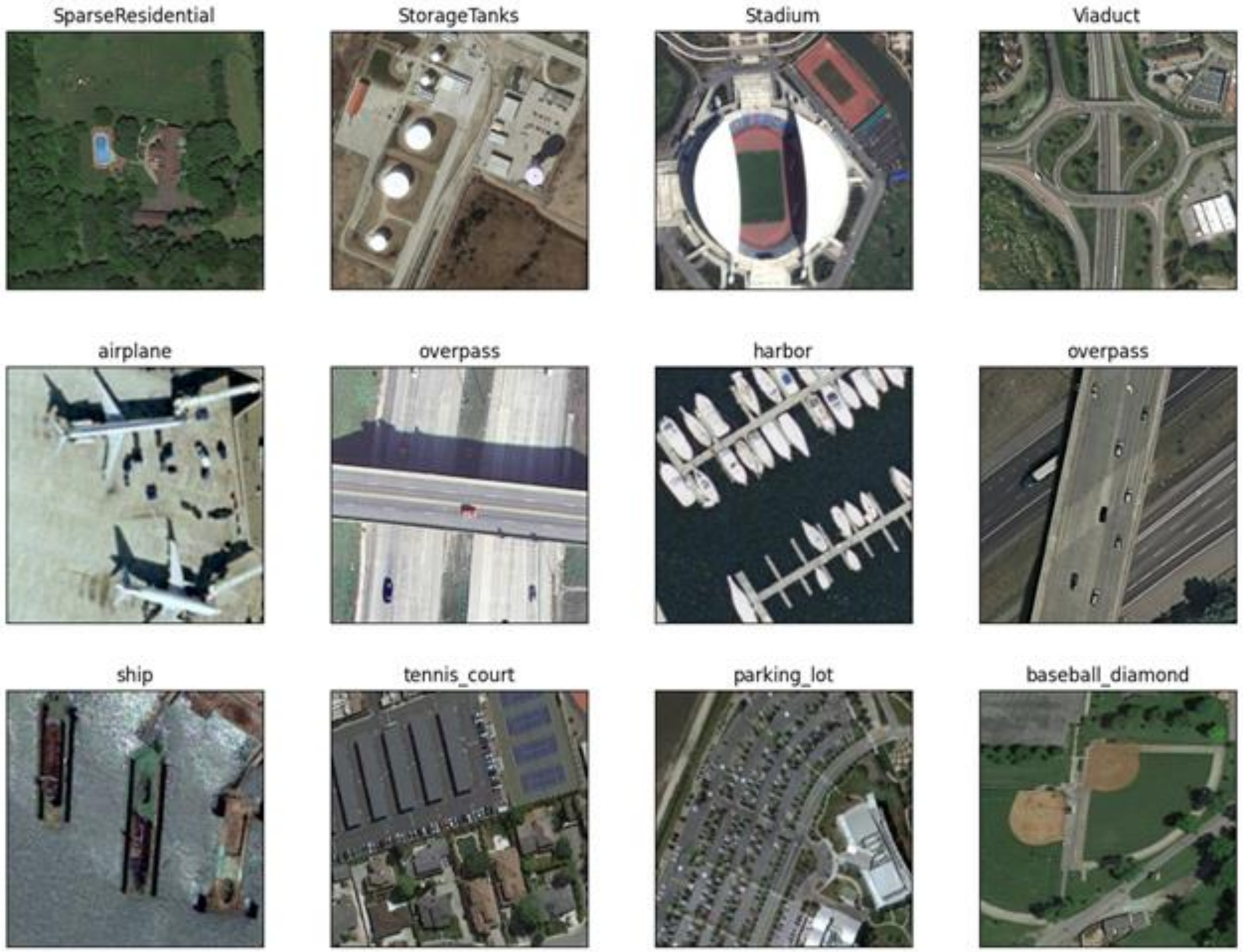
Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT at Inference Time



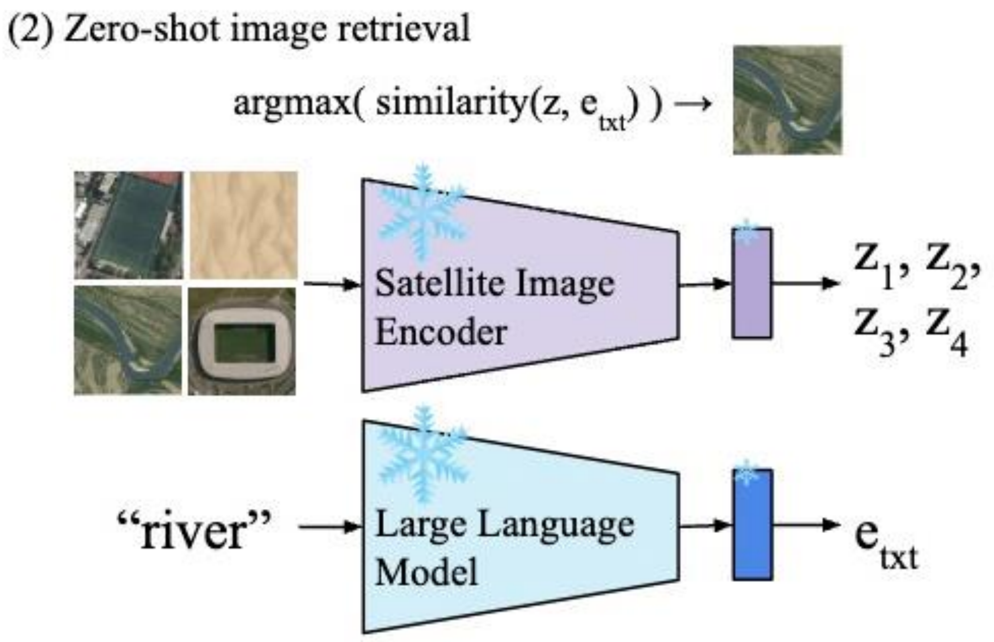
Downstream classification tasks can use the satellite representation from the encoder **without training the encoder**.

A **trainable linear layer** is then tuned for varying datasets with different classes.



Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

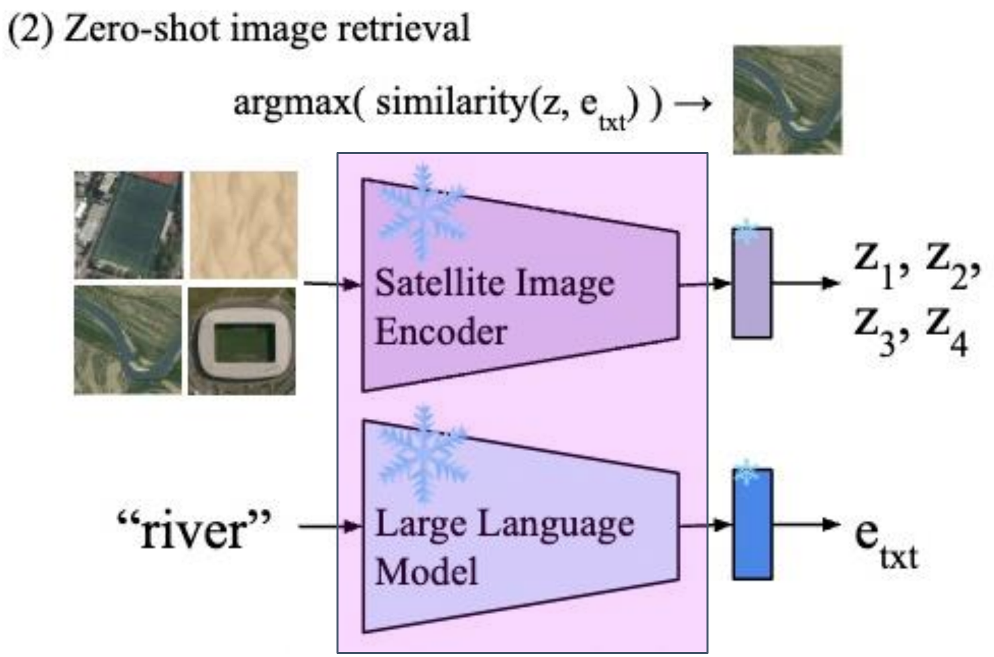
WildSAT at Inference Time



Zero-shot image retrieval requires **no additional training**.

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT at Inference Time

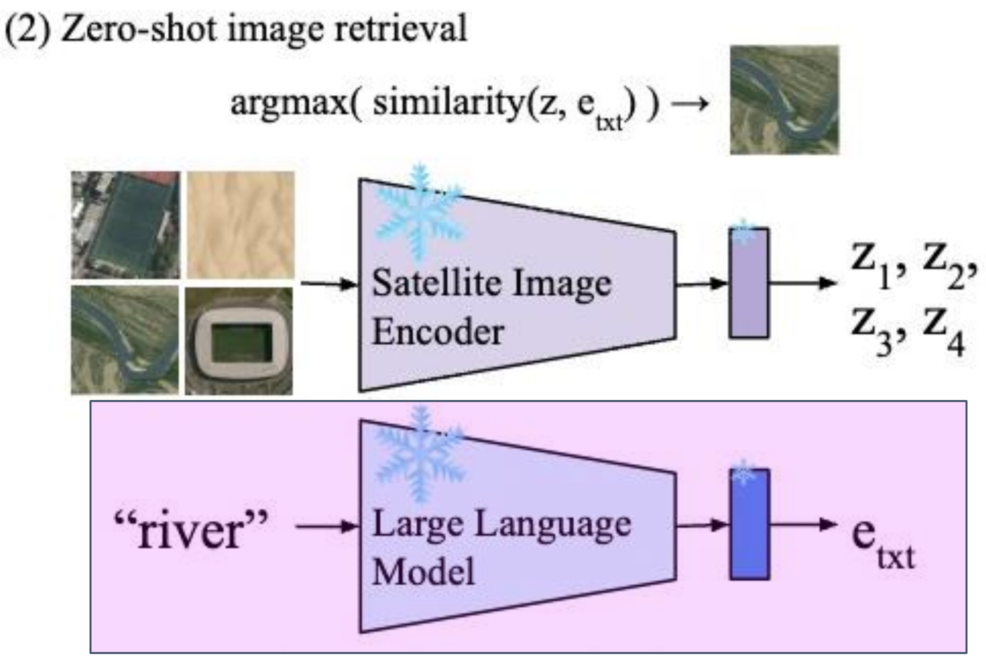


Zero-shot image retrieval requires **no additional training**.

The representation from the **image and the text encoders** are taken as is from the frozen encoders.

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT at Inference Time



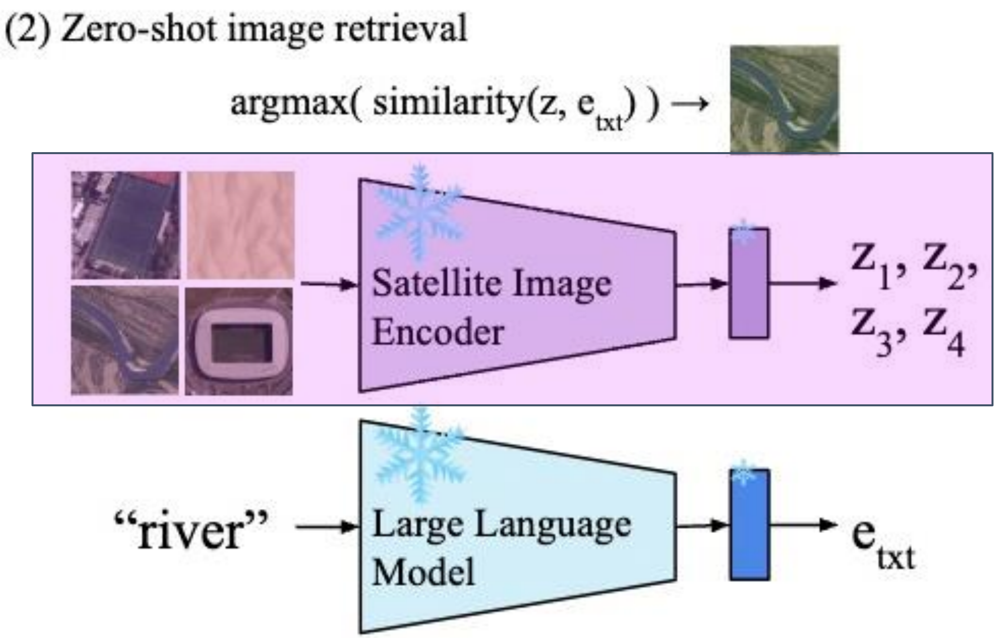
Zero-shot image retrieval requires **no additional training**.

The representation from the **image and the text encoders** are taken as is from the frozen encoders.

Given a text query (e.g., "river"), its representation/vector is computed

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT at Inference Time



Zero-shot image retrieval requires **no additional training**.

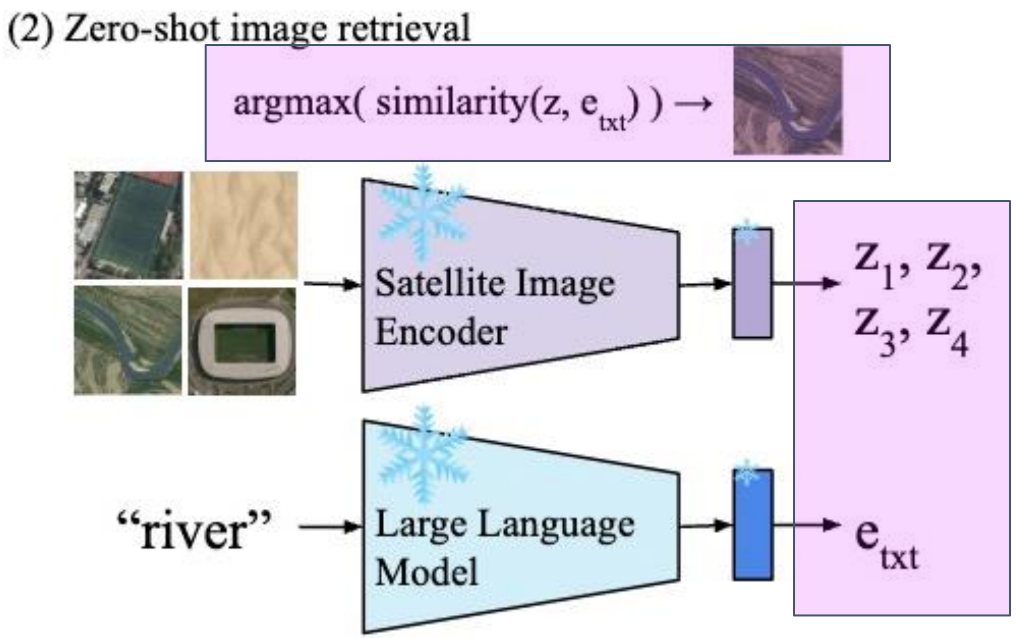
The representation from the **image and the text encoders** are taken as is from the frozen encoders.

Given a text query (e.g., "river"), its representation/vector is computed

Given a collection of images, all their corresponding representations/vectors are also computed

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

WildSAT at Inference Time



Zero-shot image retrieval requires **no additional training**.

The representation from the **image and the text encoders** are taken as is from the frozen encoders.

Given a text query (e.g., "river"), its representation/vector is computed

Given a collection of images, all their corresponding representations/vectors are also computed

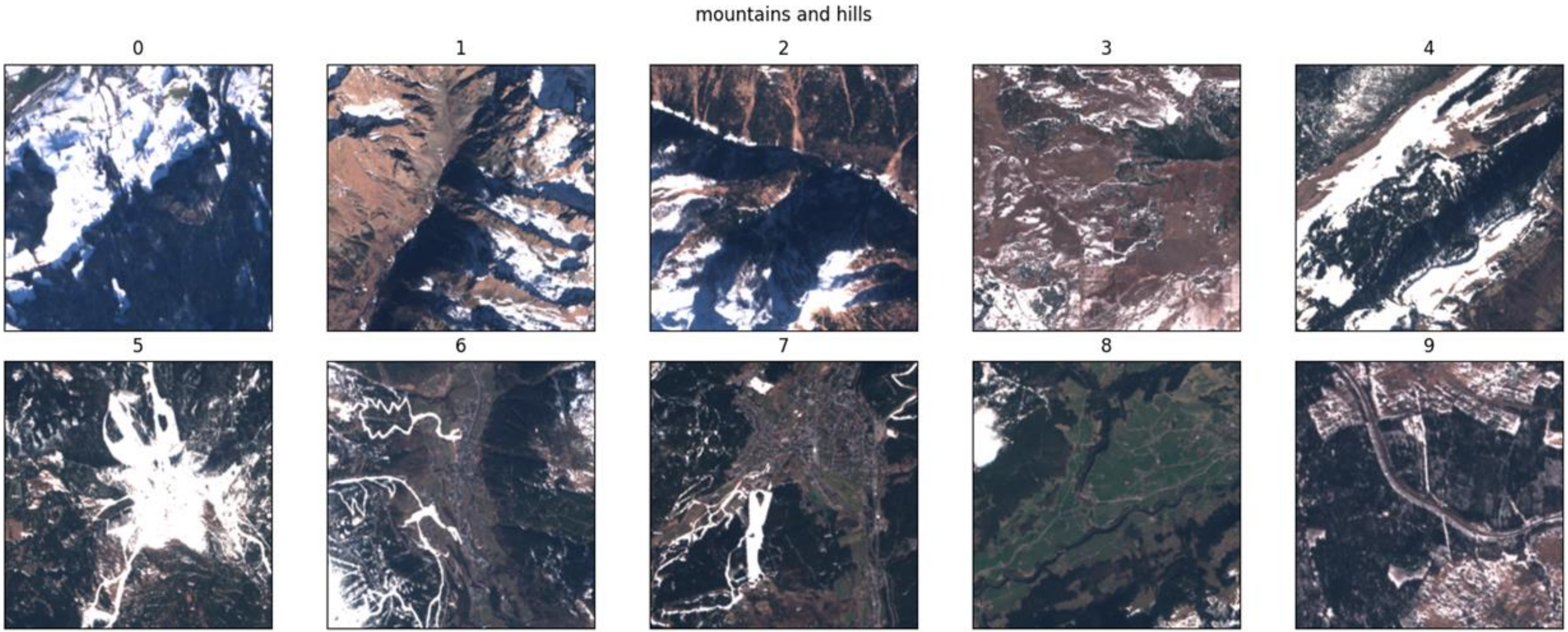
The image vector with the **highest cosine similarity** with the text vector will be returned as the output.

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

Zero-shot Retrieval Examples ("ocean")

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

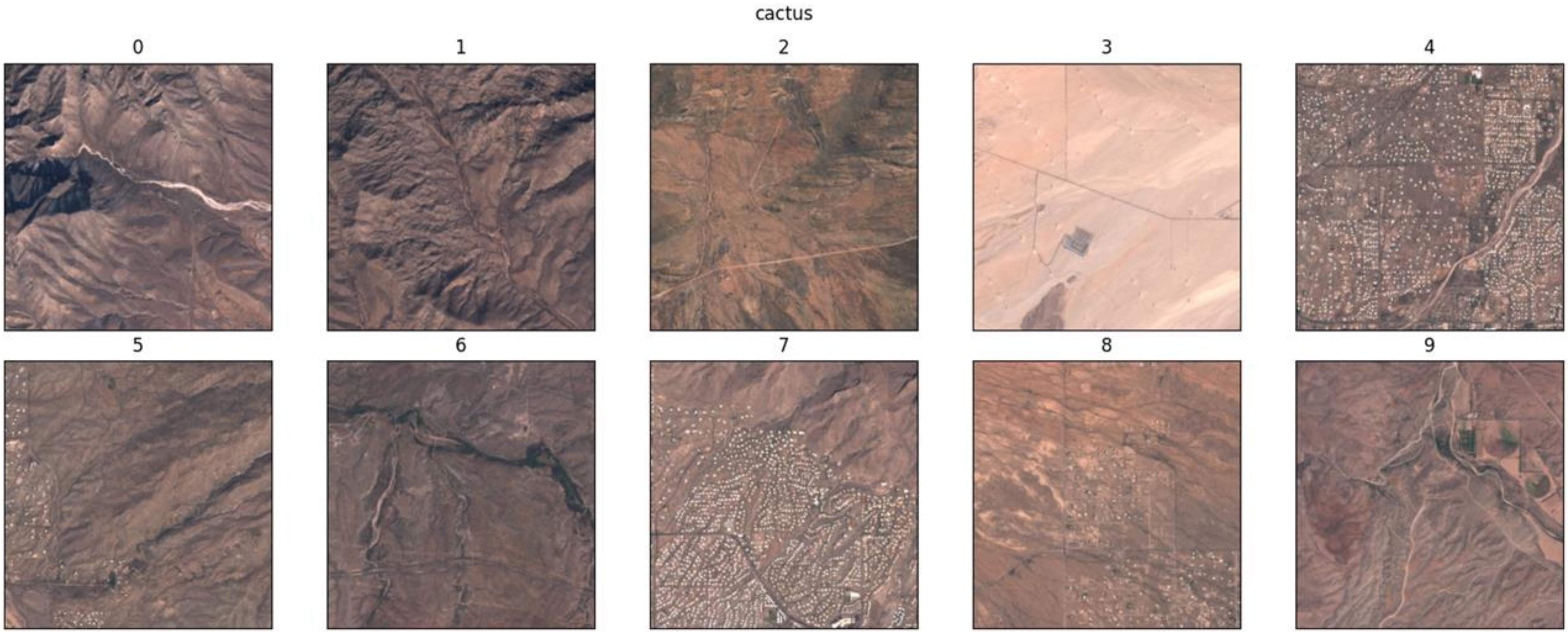
Zero-shot Retrieval Examples ("mountains and hills")



University of Massachusetts Amherst

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

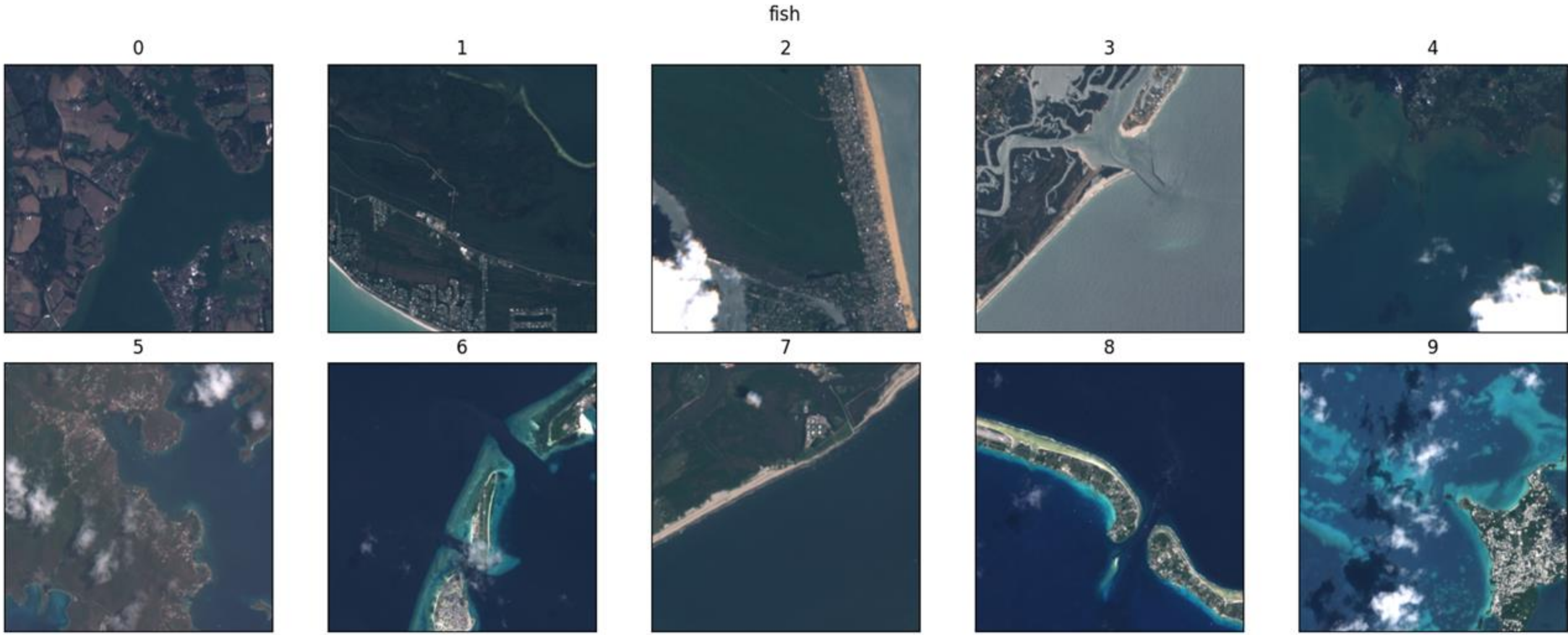
Zero-shot Retrieval Examples ("cactus")



University of Massachusetts Amherst

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

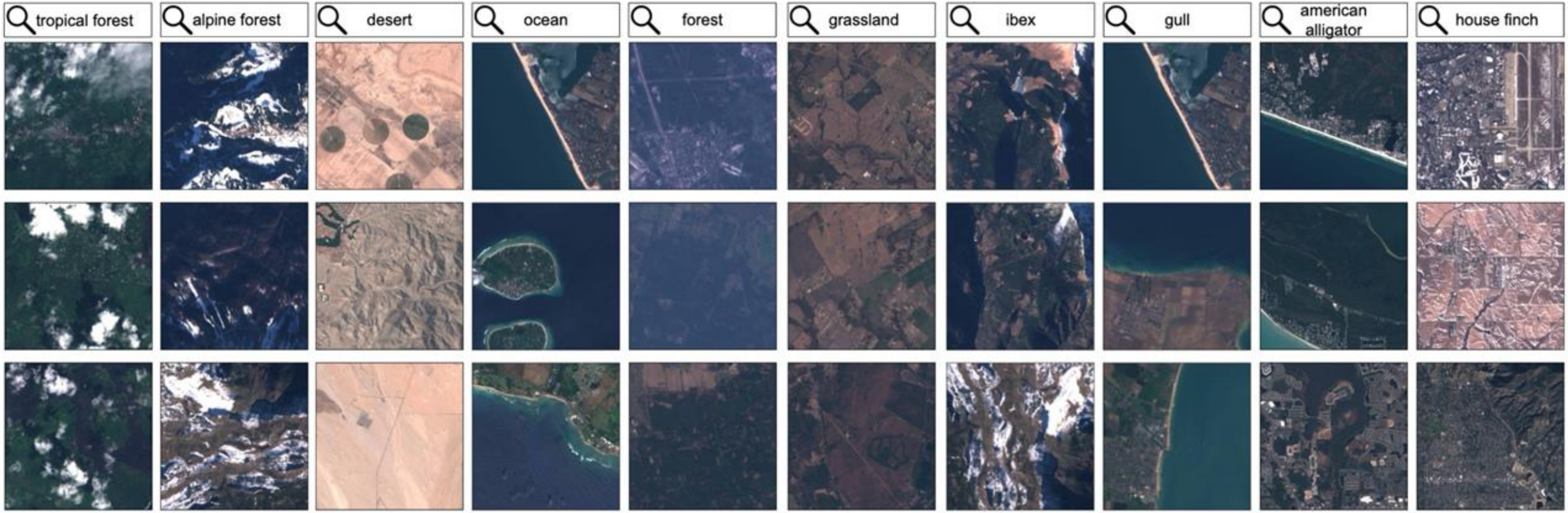
Zero-shot Retrieval Examples ("fish")



University of Massachusetts Amherst

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

More Zero-shot Satellite Image Retrieval Examples



Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

Linear probing results on downstream satellite image classification datasets

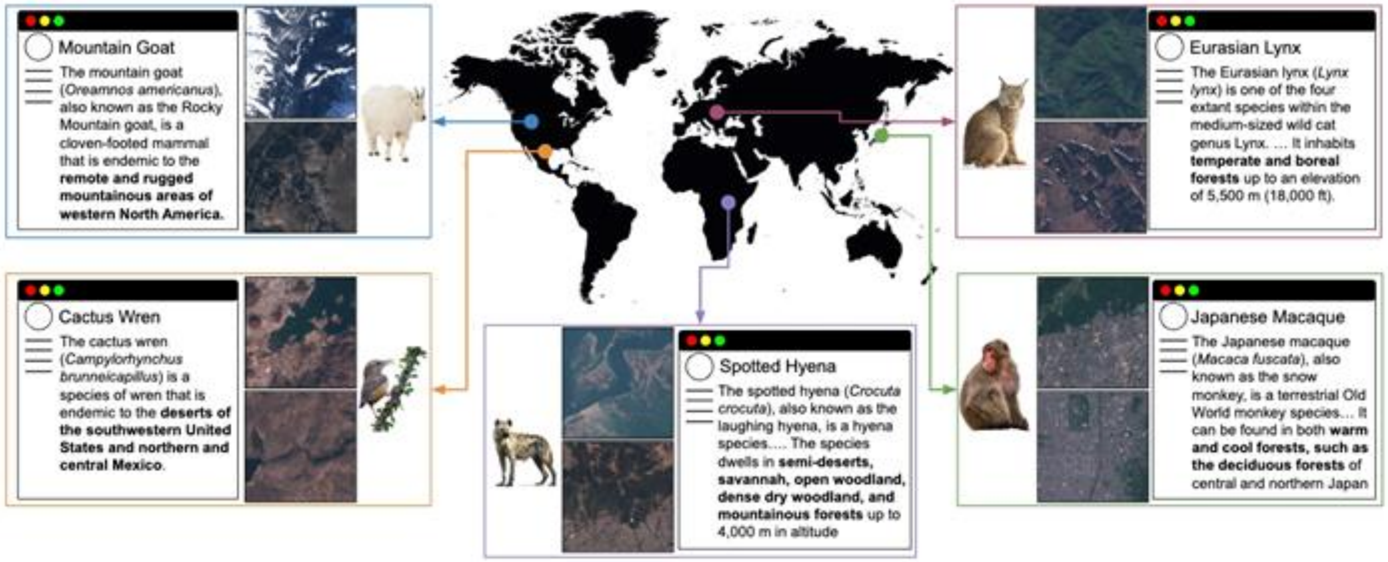
WildSAT (+WS) improves performance across different datasets and models

	Encoder	UCM [74] Base	UCM [74] +WS	AID [72] Base	AID [72] +WS	RESISC45 [7] Base	RESISC45 [7] +WS	FMoW [8] Base	FMoW [8] +WS	EuroSAT [27] Base	EuroSAT [27] +WS	So2Sat20k [78] Base	So2Sat20k [78] +WS	BEN20k [62] Base	BEN20k [62] +WS
ViT-B/16	ImageNet [13]	93.2	97.5	84.4	88.9	88.2	93.0	43.8	51.4	94.5	97.3	41.8	55.2	52.3	58.2
	MoCov3 [6]	94.2	95.1	86.0	86.9	89.1	90.3	51.1	52.9	95.9	97.1	47.6	56.6	51.6	57.0
	CLIP [57]	94.5	96.3	86.3	88.0	92.1	93.0	51.5	52.8	92.2	97.1	37.6	49.7	47.1	59.1
	Prithvi-100M* [31]	49.7	85.5	35.9	71.2	42.6	73.5	19.2	30.5	67.3	93.5	21.5	45.1	33.6	50.6
	SatCLIP* [34]	38.2	50.3	37.4	46.4	40.4	46.2	19.0	20.1	74.6	79.4	39.0	43.1	27.0	28.7
	Random weights	4.1	75.5	3.8	62.1	1.9	62.4	8.0	26.0	11.1	90.4	5.9	46.8	0.0	51.2
Swin-T	ImageNet [13]	94.0	96.9	87.9	89.0	90.4	91.8	47.6	50.7	96.2	97.3	48.3	51.5	54.1	57.7
	SatlasNet [4]	89.6	91.2	74.3	81.2	80.2	86.5	31.8	44.6	90.8	95.5	36.4	53.1	48.7	56.5
	Random weights	21.0	81.7	19.5	72.0	19.9	74.9	12.1	33.4	59.9	92.7	21.9	45.9	9.8	52.4
ResNet50	ImageNet [13]	94.2	93.6	87.8	86.7	90.5	90.1	47.3	46.0	95.5	96.0	36.1	46.6	55.8	57.5
	MoCov3 [6]	92.0	93.5	83.0	83.3	88.0	87.6	50.2	45.7	93.5	95.1	27.2	42.5	46.6	53.8
	SatlasNet [4]	86.8	90.1	72.5	79.4	81.8	85.4	34.7	42.4	93.5	95.4	33.9	44.8	44.9	56.4
	SeCo [46]	86.1	88.8	74.3	79.6	80.2	86.3	35.9	42.8	89.7	95.5	39.9	46.0	44.3	57.3
	SatCLIP* [34]	69.4	76.2	63.1	71.8	70.2	78.8	36.2	39.9	83.4	92.9	45.4	44.9	42.3	48.2
	Random weights	24.7	79.9	22.3	68.2	24.5	74.7	12.7	36.9	65.2	92.2	5.9	42.3	19.9	51.3
Overall average		68.8	**86.1**	61.2	**77.0**	65.3	**81.0**	33.4	**41.1**	80.2	**93.8**	32.6	**47.6**	38.5	**53.1**
Average w/o random		81.8	**87.9**	72.7	**79.4**	77.8	**83.5**	39.0	**43.3**	88.9	**94.3**	37.9	**48.3**	45.7	**53.4**

Conclusion

WildSAT: Learning Satellite Image Representations from Wildlife Observations

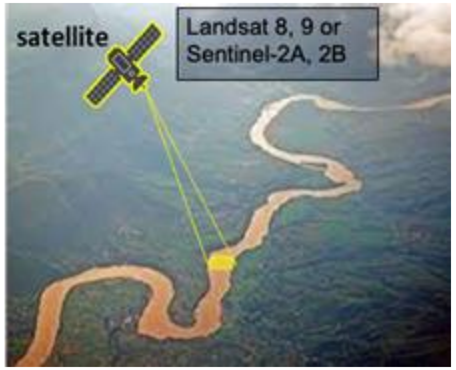
- We showed that information on habitats of wildlife can be used to improve satellite image representations in ML models
  - Can improve performance on downstream satellite image tasks
  - Can be used for zero-shot image retrieval using arbitrary text input



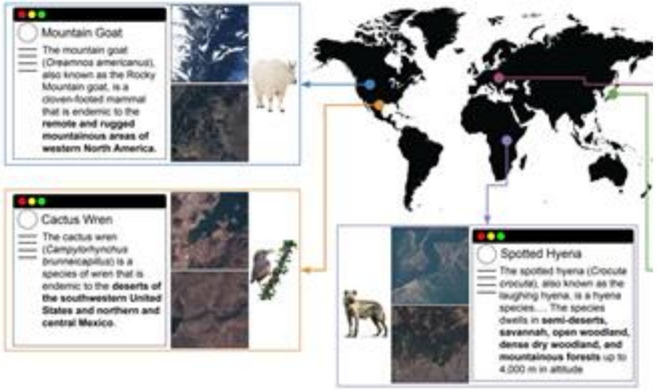
arxiv.org/abs/2412.14428

Computer Vision in the Wild

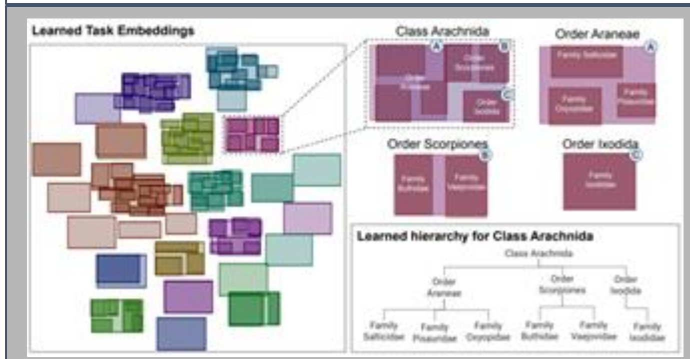
- In this talk, we will explore domain-specific applications of computer vision



1. Remote monitoring of global water quality using satellite images



2. Helping wildlife conservation efforts through wildlife habitat information and satellite imagery



3. Modeling relationships between domain-specific tasks for efficient ML model training

UMassAmherst | Manning College of Information & Computer Sciences

CVPR SEATTLE, WA JUNE 17-21, 2024

Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships

Rangel Daroya, Aaron Sun, Subhransu Maji

University of Massachusetts Amherst

Overview: Transfer Learning

Training #1 (Pre-training)

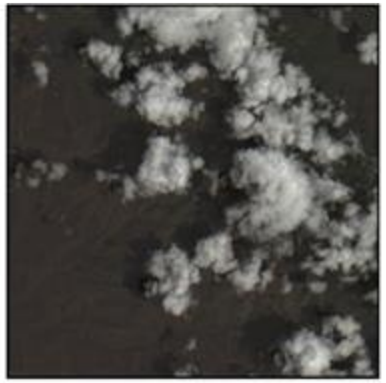
Data

Label

Machine Learning Model

"dog"

Training #2 (Fine-tuning)

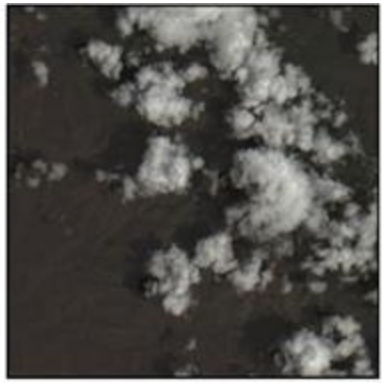


Machine Learning Model

Overview: Transfer Learning

Training #1 (Pre-training)

Data



Machine Learning Model

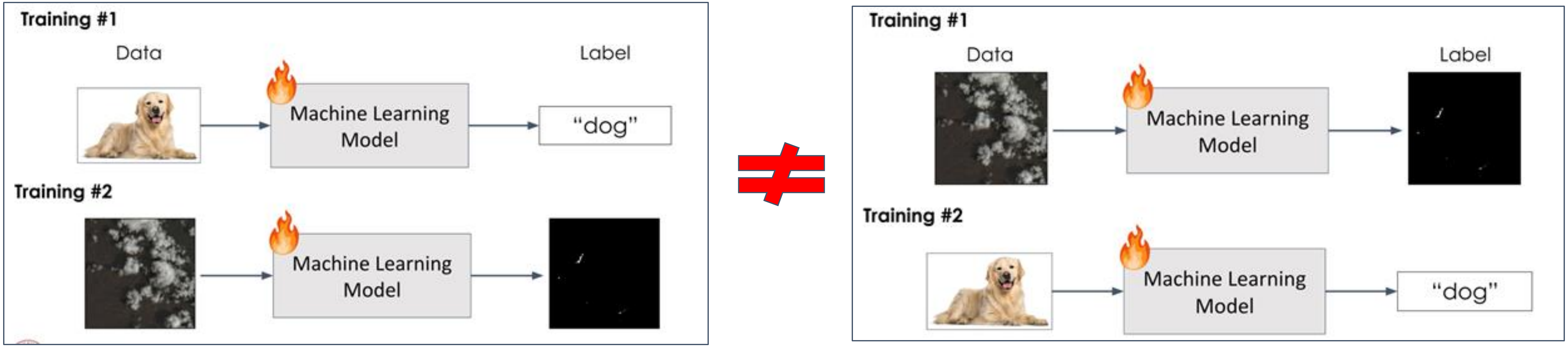
Label

Training #2 (Fine-tuning)

Machine Learning Model

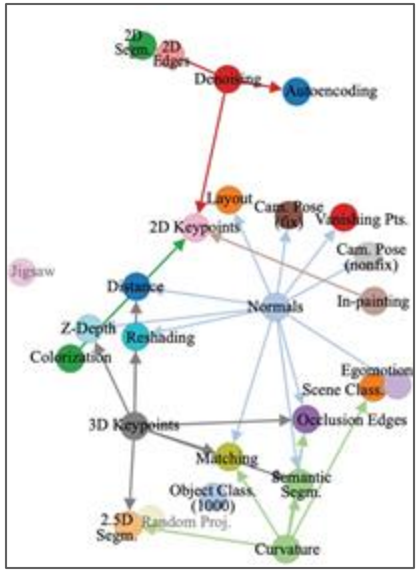
“dog”

Asymmetry of transfer learning

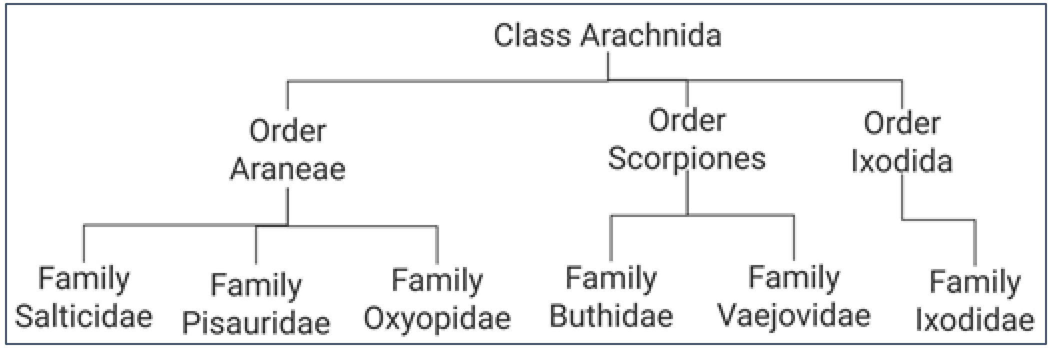


Problem Overview

- Modeling and visualizing relationships between tasks or datasets is important for solving various meta-tasks
  - Dataset Discovery, Multitask Learning, Transfer Learning
- However, many relationships are asymmetric (e.g., containment, transferability)



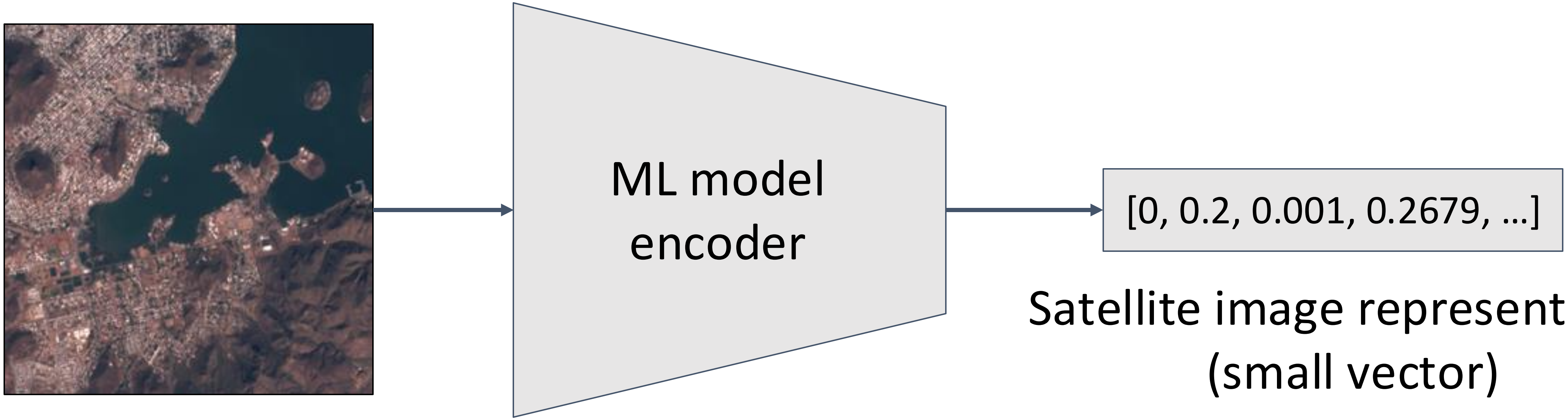
(a) Transfer Learning task affinity (Taskonomy [1])



(b) Taxonomy (iNaturalist + CUB)

Recall: Satellite image representations

Satellite image representations refer to the encoded **satellite image** using a given ML model



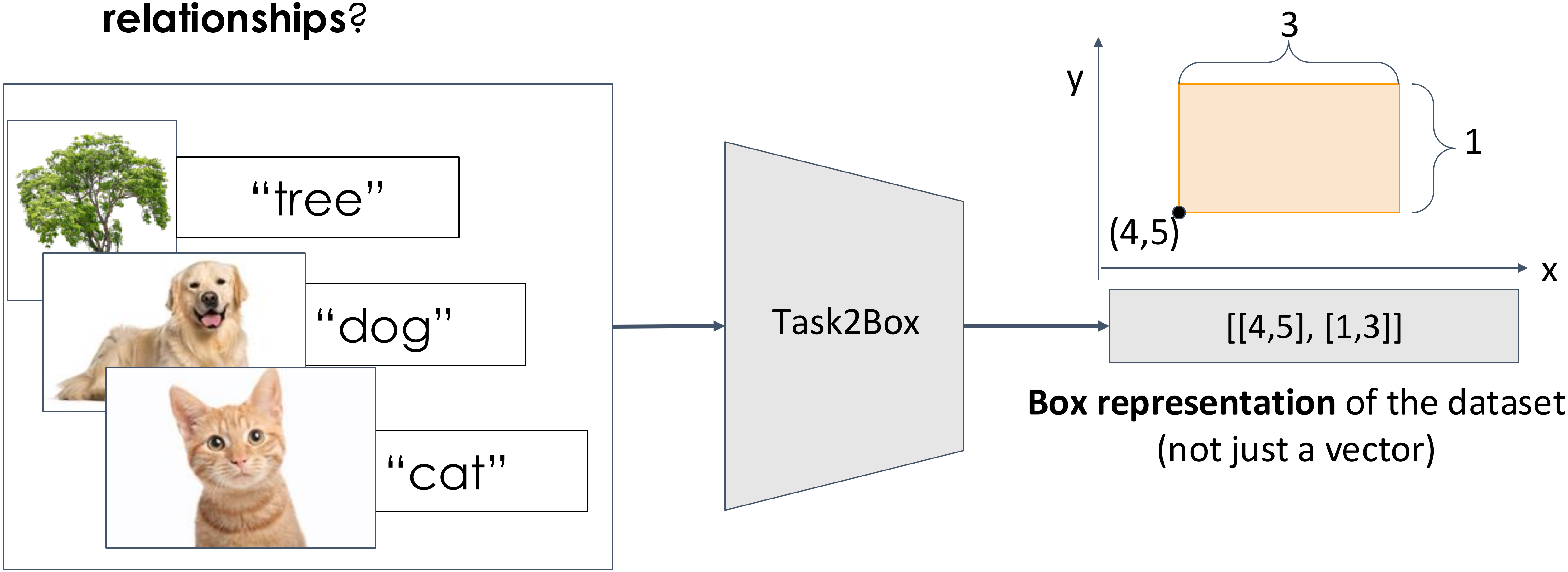
Satellite image representation
(small vector)

Satellite image

Rangel Daroya, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.

Task Representations

- Can we create **task representations** that preserve **asymmetric relationships**?

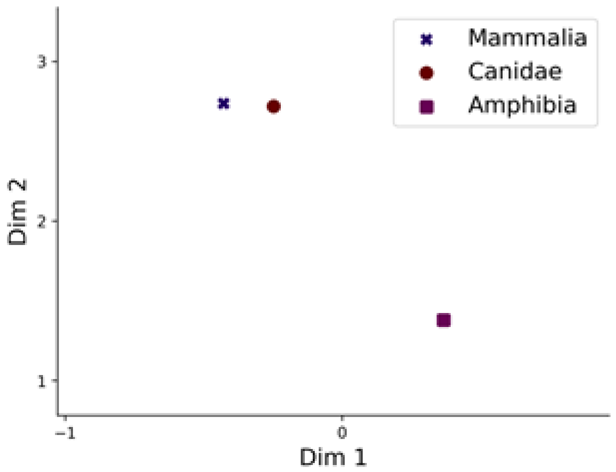


Dataset

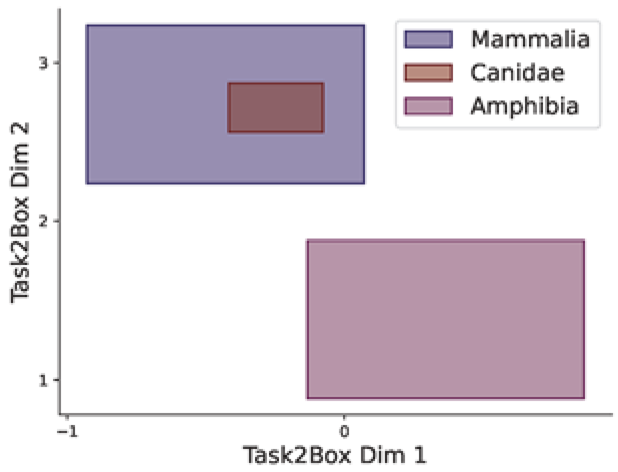
Problem Overview

- Can we create **task representations** that preserve **asymmetric relationships**?
- Proposed Solution:
  - Use a model to learn box embeddings (axis-aligned hyperrectangles) to represent each dataset in a low dimension

Tasks as points



Tasks as boxes

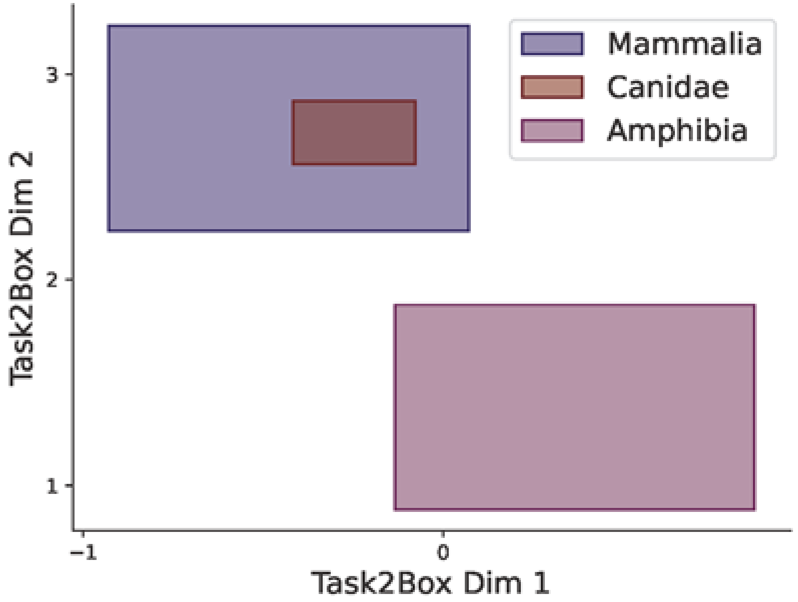


University of Massachusetts Amherst

Why boxes?

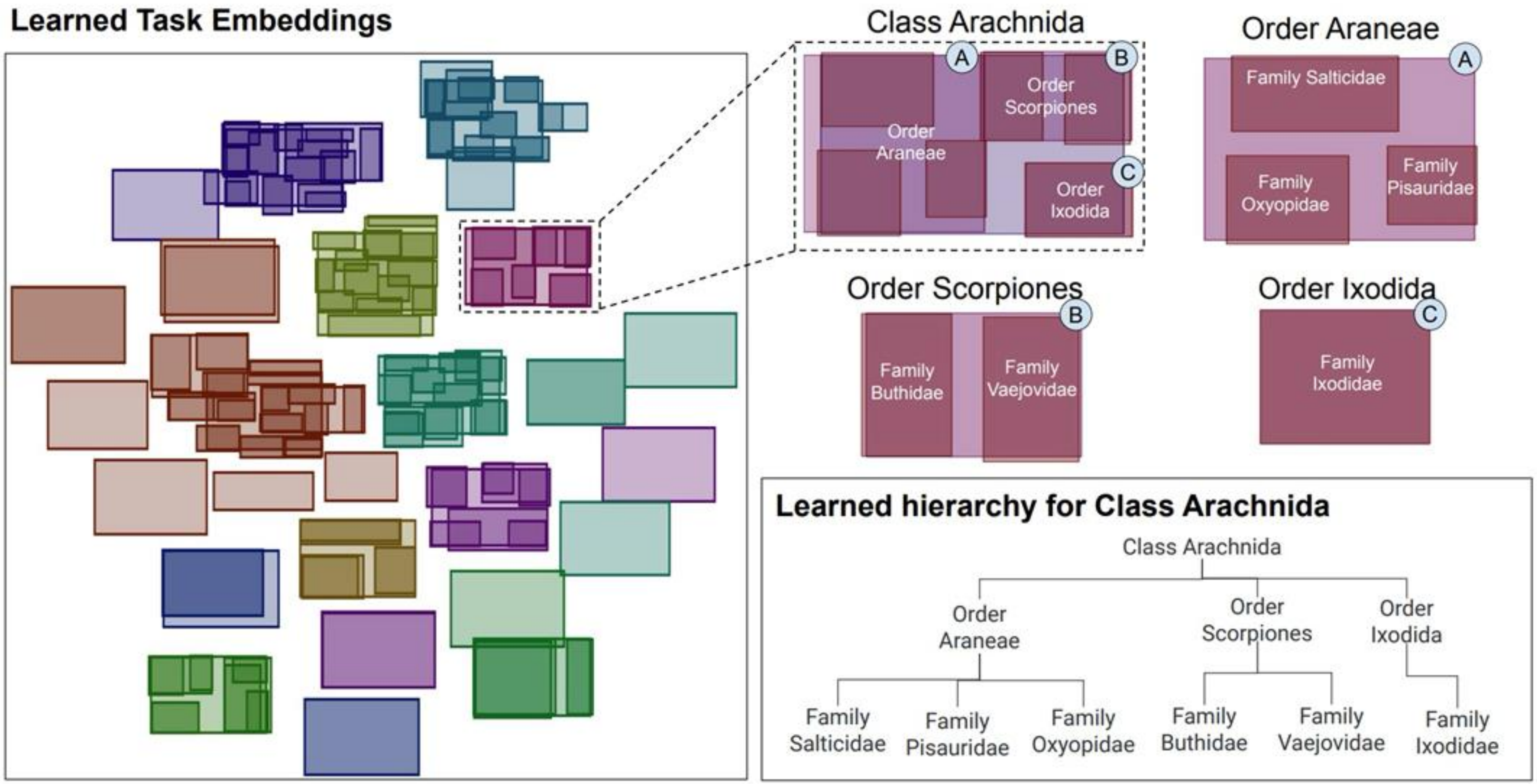
- Boxes can represent asymmetric relationships (unlike points)
  - hierarchy, transferability
- It can easily be visualized and interpreted
- Boxes are closed under intersection
  - I.e., the intersection of two boxes will always be a box

Tasks as boxes



University of Massachusetts Amherst

Task2Box Accurately Models Hierarchical Relationships



University of Massachusetts Amherst

Rangel Daroya, Aaron Sun, Subhransu Maji. "Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 28827-28837

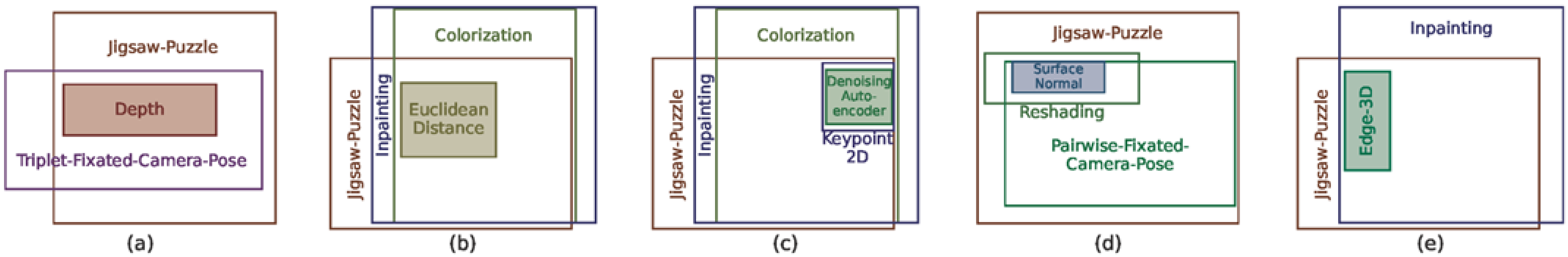
Task2Box Accurately Models Hierarchical Relationships

Method	Existing Datasets			Novel Datasets		
	μ_{CLIP}	$[\mu, \sigma^2]_{CLIP}$	FIM	μ_{CLIP}	$[\mu, \sigma^2]_{CLIP}$	FIM
TASK2BOX (2D)	69.23%	67.84%	39.61%	50.07%	39.66%	10.06%
TASK2BOX (3D)	<u>79.66%</u>	<u>79.35%</u>	<u>57.63%</u>	<u>70.04%</u>	<u>64.53%</u>	<u>20.65%</u>
TASK2BOX (5D)	**84.67%**	**82.41%**	**79.72%**	**73.79%**	**72.11%**	**34.88%**
MLP Classifier	45.25%	61.45%	26.34%	39.06%	44.54%	19.90%
Linear Classifier	4.40%	3.11%	7.06%	4.77%	5.87%	15.92%
KL Divergence	-	6.58%	7.94%	-	5.90%	0.00%
Asymmetric Cosine	9.29%	11.54%	2.83%	1.47%	1.47%	1.47%
Asymmetric Euclidean	1.71%	1.71%	8.53%	1.47%	1.47%	1.91%
Random		2.06%			1.49%	

University of Massachusetts Amherst

Rangel Daroya, Aaron Sun, Subhransu Maji. "Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 28827-28837

Task2Box Accurately Models Transfer Learning Relationships



- The figure shows predicted source tasks (larger boxes) that transfer well to target tasks (smaller shaded boxes).
- Task2Box can generalize on task affinity values from Taskonomy to predict and show transferability between tasks.

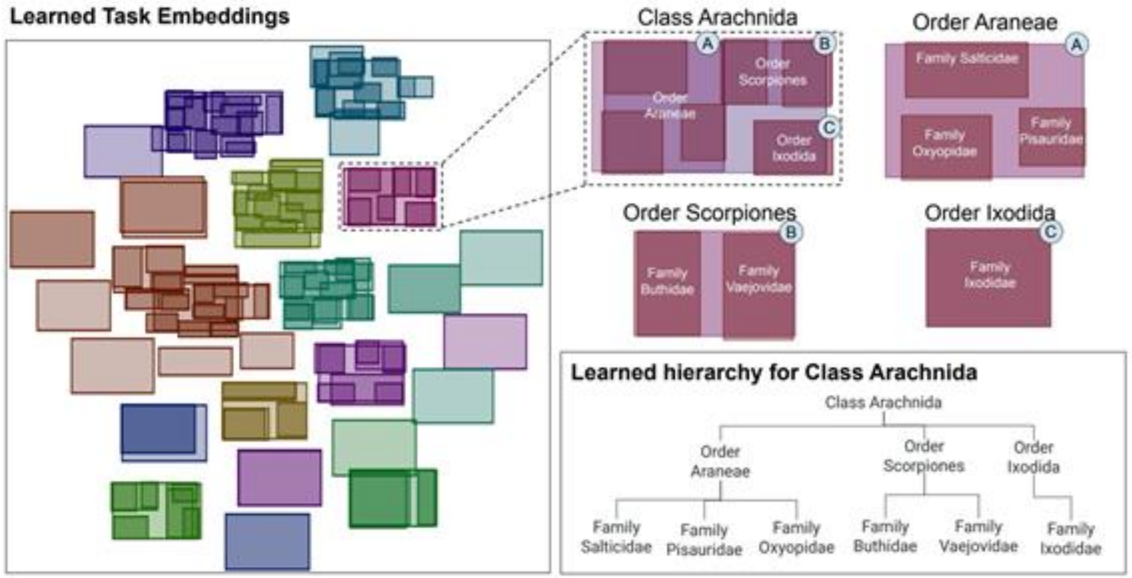
Method	**Existing Datasets** Spearman's ρ	**Novel Datasets** Spearman's ρ
TASK2BOX (2D)	0.85 ± 0.06	0.12 ± 0.21
TASK2BOX (3D)	<u>0.93 ± 0.02</u>	$\mathbf{0.48 \pm 0.24}$
TASK2BOX (5D)	$\mathbf{0.94 \pm 0.03}$	0.39 ± 0.22
MLP	0.88 ± 0.06	0.31 ± 0.18
Linear	0.75 ± 0.11	<u>0.40 ± 0.24</u>
Random	0.05 ± 0.14	0.15 ± 0.07

Rangel Daroya, Aaron Sun, Subhransu Maji. "Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 28827-28837

Conclusion

Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships

- Proposed a method of representing tasks as **box embeddings**
- The representations are **interpretable** with low dimensionality
- Shows that **hierarchical** and **transfer learning relationships** can be accurately modeled

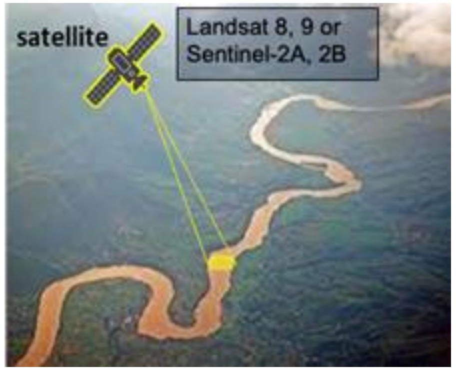


github.com/cvl-umass/task2box

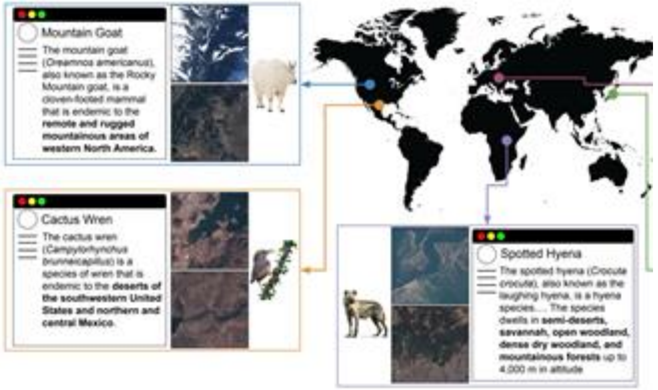
Rangel Daroya, Aaron Sun, Subhransu Maji. "Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 28827-28837

Computer Vision in the Wild

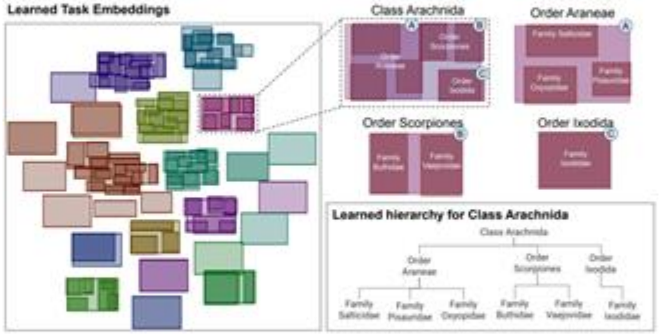
- In this talk, we will explore domain-specific applications of computer vision



1. Remote monitoring of global water quality using satellite images



2. Helping wildlife conservation efforts through wildlife habitat information and satellite imagery



3. Modeling relationships between domain-specific tasks for efficient ML model training

Thank you to all collaborators and colleagues

.. and many more

Thank you!

Papers in this talk:

1. **Rangel Daroya**, Luisa Vieira Lucchese, Travis Simmons, Punwath Prum, Tamlin Pavelsky, John Gardner, Colin Gleason, Subhransu Maji. "Improving Satellite Imagery Masking using Multi-task and Transfer Learning", in *arXiv preprint arXiv:2412. 08545*, 2024.
2. **Rangel Daroya**, Elijah Cole, Oisin Mac Aodha, Grant Van Horn, Subhransu Maji. "WildSAT: Learning Satellite Image Representations from Wildlife Observations," in *arXiv preprint arXiv:2412.14428*, 2024.
3. **Rangel Daroya**, Aaron Sun, Subhransu Maji. "Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships," Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 28827-28837

Contact

Rangel Daroya

Email: rdaroya@umass.edu

rangeldaroya.github.io

University of Massachusetts Amherst