

The logo for the 2024 Embedded VISION Summit is centered within a white octagonal shape. The octagon is surrounded by a colorful, multi-layered border of overlapping geometric shapes in shades of purple, blue, green, yellow, and orange. The text inside the octagon reads "2024 embedded VISION SUMMIT®".

2024
embedded
VISION
SUMMIT®

Data-Efficient & Generalizable: The Domain-Specific Small Vision Model Revolution

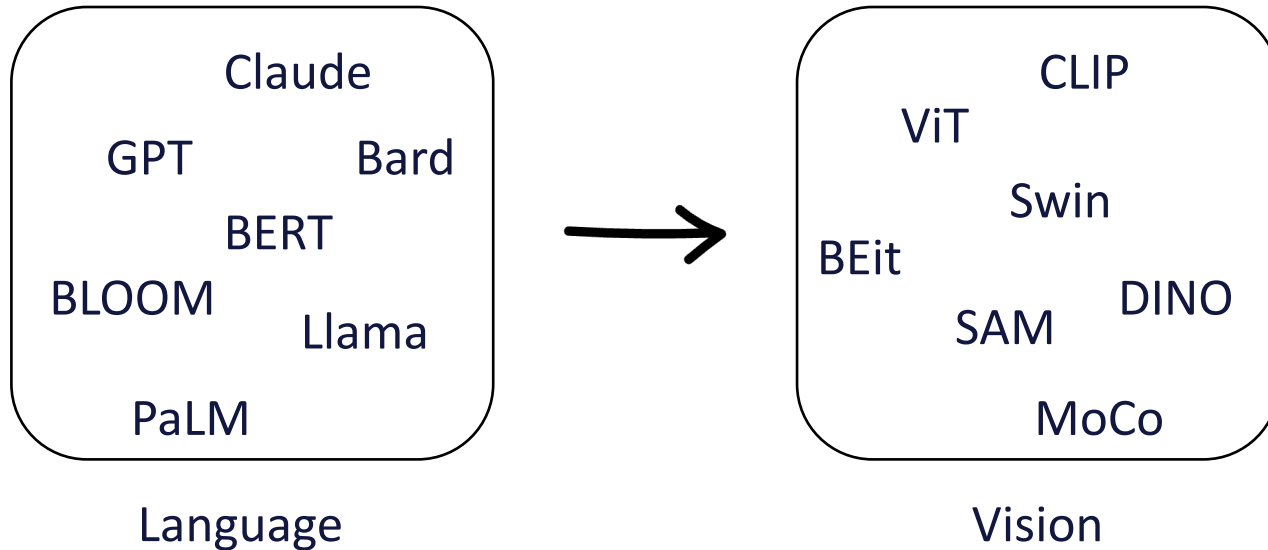
Heather D. Couture

Founder & Computer Vision Consultant

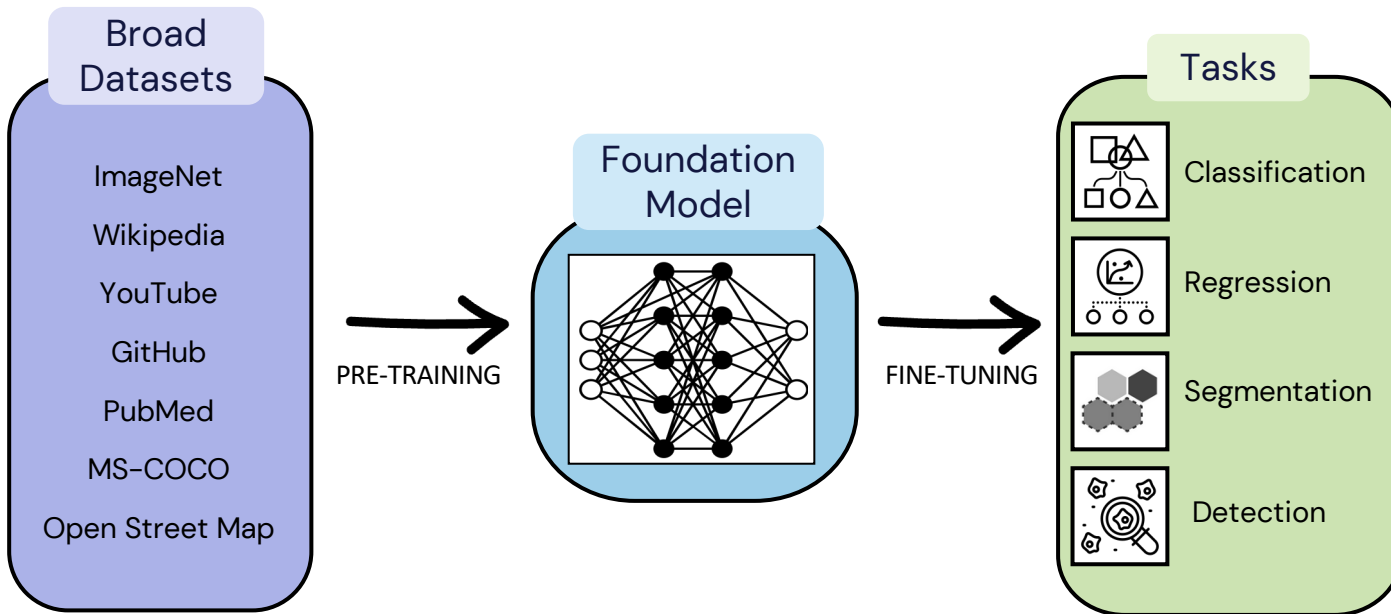
Pixel Scientia Labs



From Large Language Models to Large Vision Models



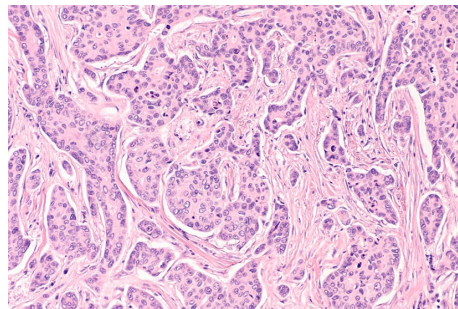
Foundation Models: Generality & Adaptability



Problem: Unique Imaging Modality

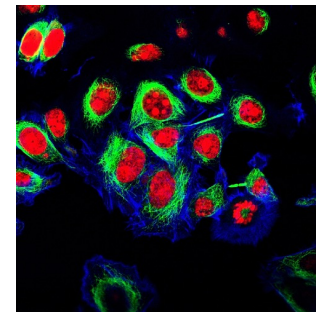


ImageNet



Histopathology

Image credit: Shutterstock



Fluorescence Microscopy

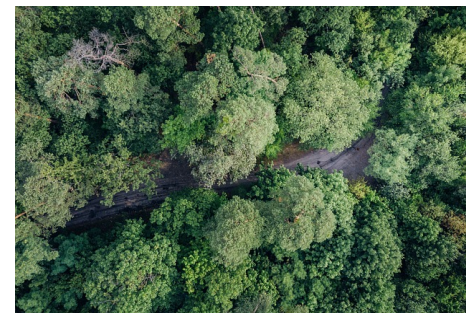
Image credit: Shutterstock

VS



Multispectral Satellite

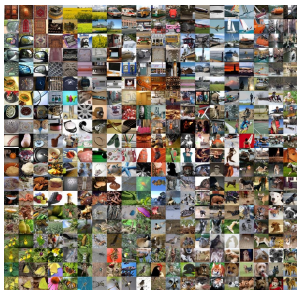
Image credit: ESA



Drone

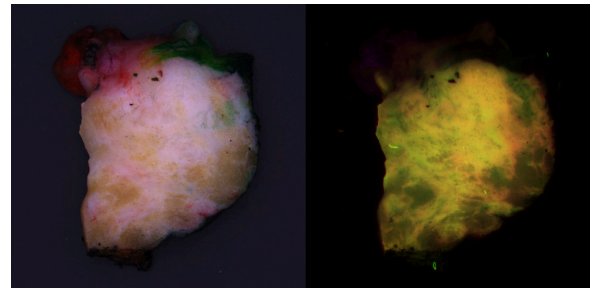
Image credit: Pixabay

Problem: Limited Data



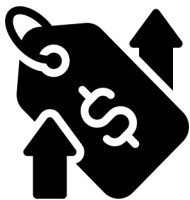
ImageNet

1.2 million vs. 200 images



New medical imaging device

Data collection and labeling can be...



Expensive



Time-consuming



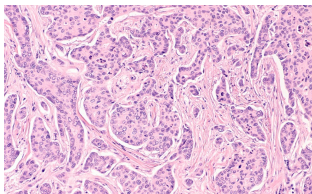
Difficult

Problem: Compute Resource Constraints

Model size	Model name	# parameters (million)	FLOPS per inference (billion)
Small	MobileNetV2	7	1.2
	ResNet18	12	1.8
	ResNet50	26	4.1
	ViT-Small	22	4.6
	Swin-Tiny	28	4.5
Medium	ResNet101	45	7.6
	Swin-Small	50	8.7
	ViT-Base	87	17.6
	Swin-Base	88	15.5
Large	Swin-Large	197	34.5
	ViT-Large	304	61.6
	ViT-Giant	1843	2860

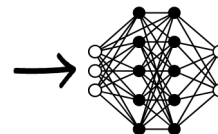
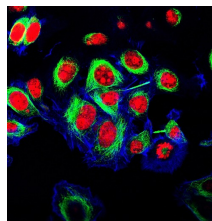
Publicly-available foundation models are getting larger

Solution: *Domain-Specific* Foundation Models



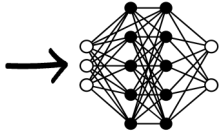
Histopathology

Image credit: Shutterstock



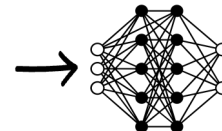
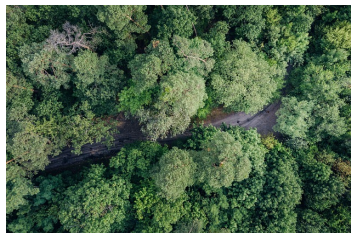
Fluorescence Microscopy

Image credit: Shutterstock



Multispectral Satellite

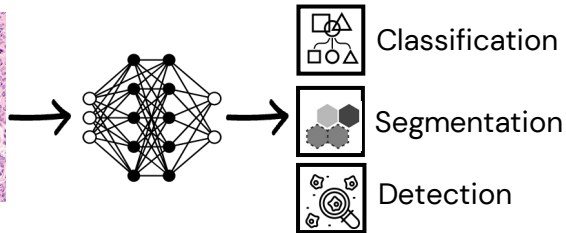
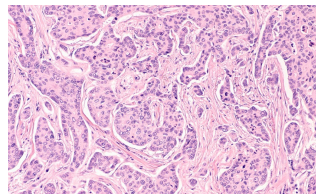
Image credit: ESA



Forestry Drone

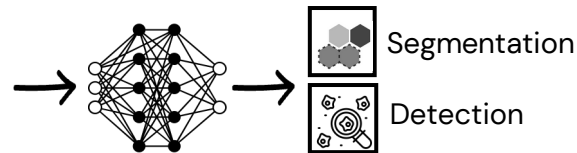
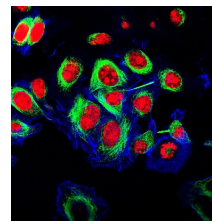
Image credit: Pixabay

Solution: *Domain-Specific* Foundation Models



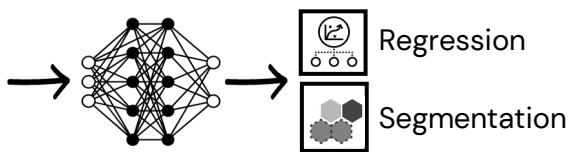
Histopathology

Image credit: Shutterstock



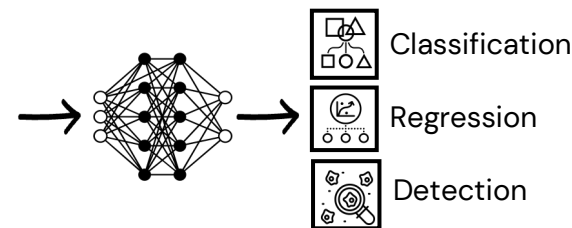
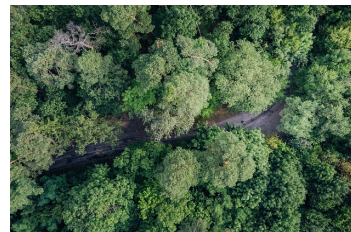
Fluorescence Microscopy

Image credit: Shutterstock



Multispectral Satellite

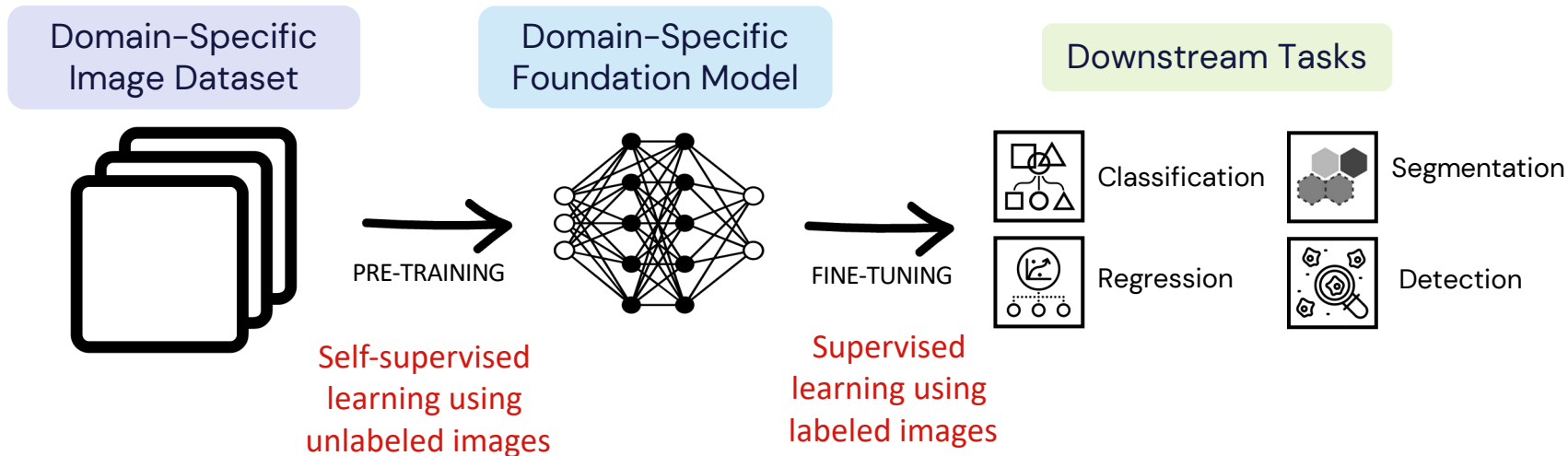
Image credit: ESA



Forestry Drone

Image credit: Pixabay

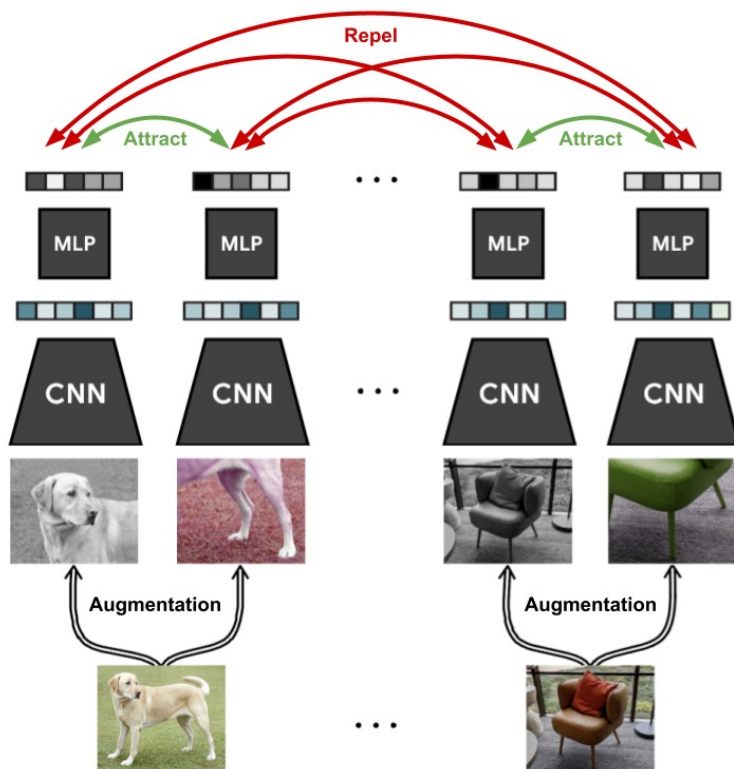
Pre-Training and Fine-Tuning



- Most computationally-intensive step
- Done once
- Select a smaller architecture for improved computation speed (training and inference)
- Uses a pretext task instead of manual labels

- For each downstream task
- Less computation
- Could be as simple as linear model with no fine-tuning

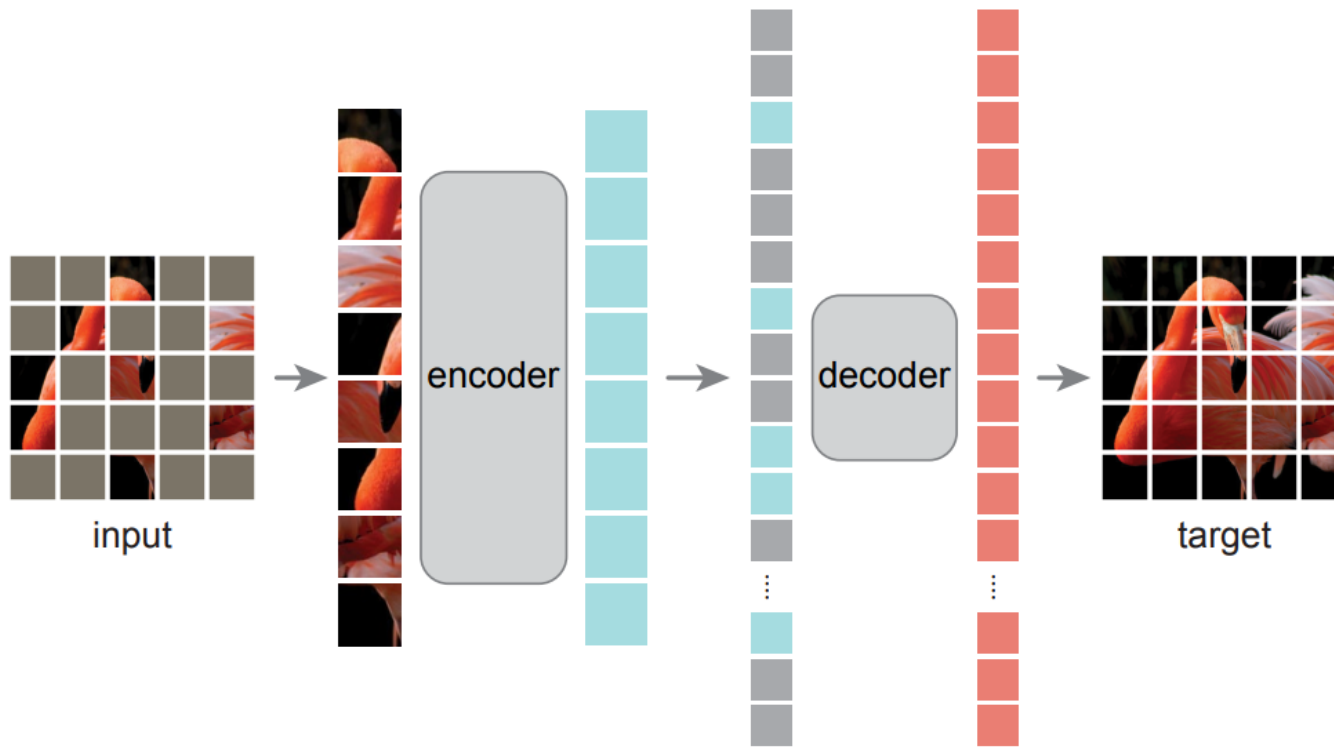
Self-Supervised Pretext Task: Contrastive



No manual labels needed

Source: <https://blog.research.google/2020/04/advancing-self-supervised-and-semi.html>

Self-Supervised Pretext Task: Masked Autoencoder



Source: He, Masked Autoencoders Are Scalable Vision Learners, 2021

Example 1: Satellite

EuroSAT: land cover classification, 27k images, 80/20 pre-train/test



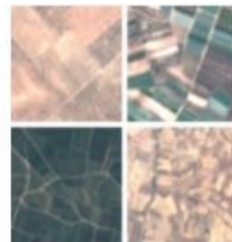
(a) Industrial Buildings



(b) Residential Buildings



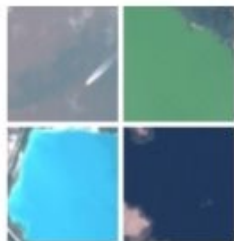
(c) Annual Crop



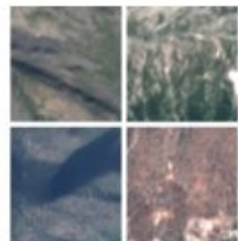
(d) Permanent Crop



(e) River



(f) Sea & Lake



(g) Herbaceous Vegetation



(h) Highway



(i) Pasture

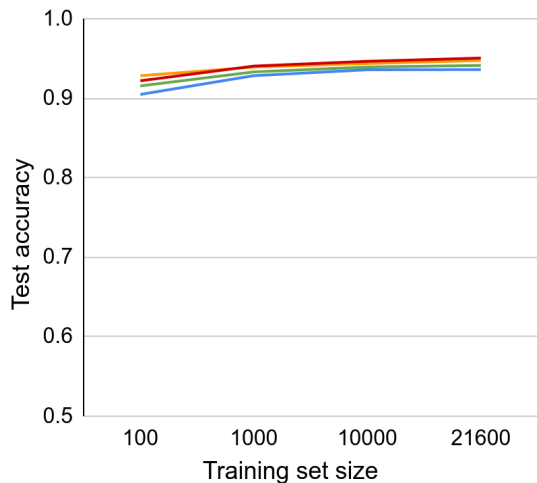


(j) Forest

Example 1: Satellite

Pre-Training on EuroSAT

Pretext Task

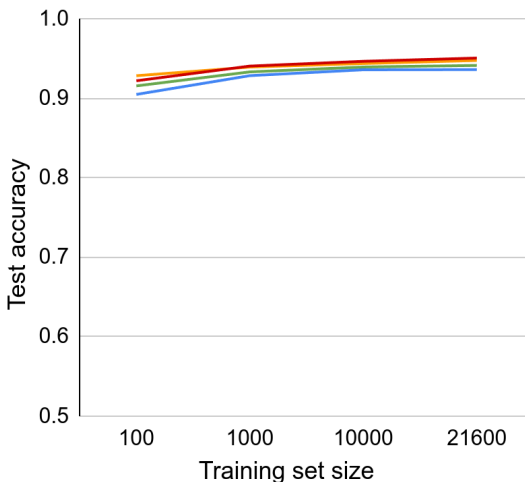


Little difference

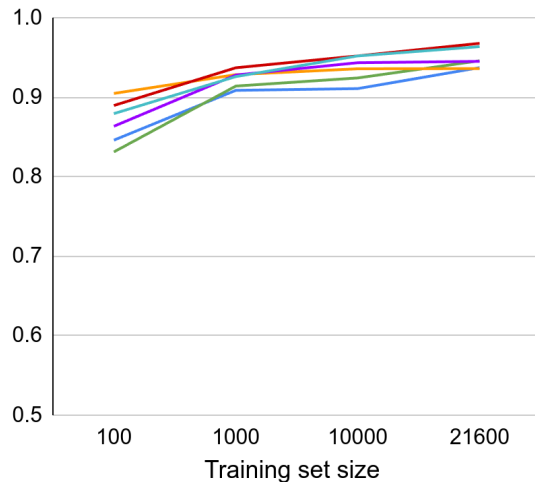
Example 1: Satellite

Pre-Training on EuroSAT

Pretext Task



Model Size



— SimCLR ResNet18 — MoCoV2 ResNet18
— SimSiam ResNet18 — VicReg ResNet18

— SimCLR ViT-T (5m) — SimCLR ViT-T (11m)
— SimCLR ResNet18 (12m) — SimCLR ViT-S (22m)
— SimCLR ResNet50 (26m) — SimCLR ViT-B (88m)

Little difference

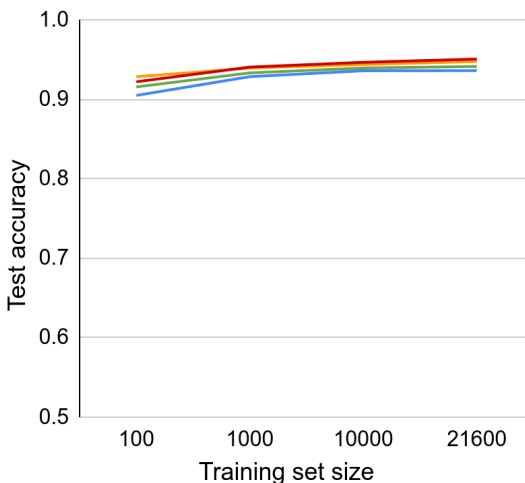
Smaller models sufficient for small training sets
Large models best for large training sets

Example 1: Satellite

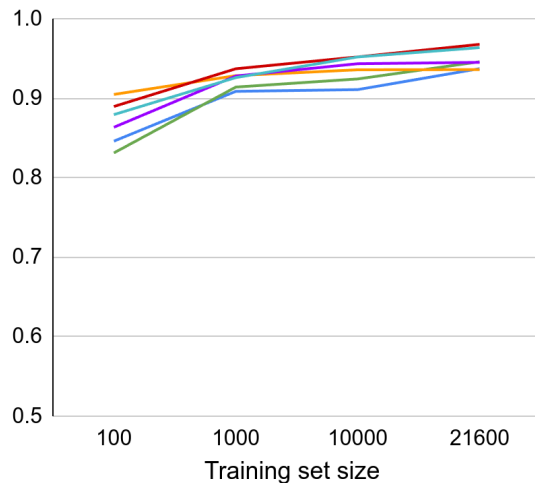
Small domain-specific models are superior for small training sets

Pre-Training on EuroSAT

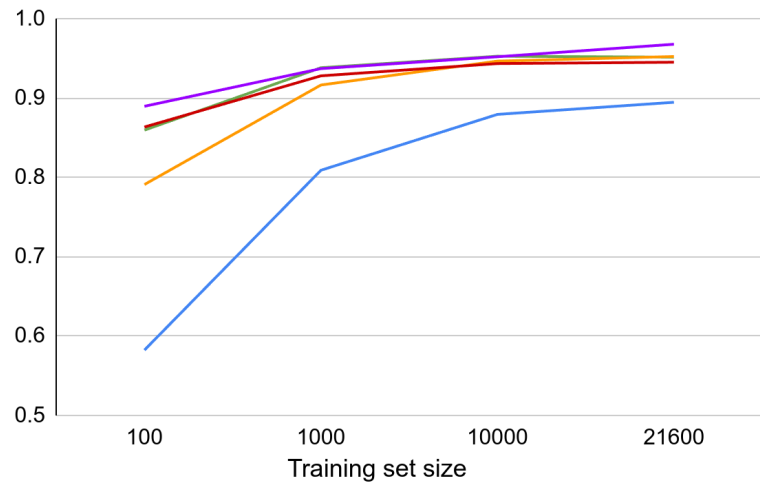
Pretext Task



Model Size



Pre-Training on Different Datasets



- SimCLR ResNet18
- SimSiam ResNet18
- MoCoV2 ResNet18
- VicReg ResNet18

- SimCLR ViT-T (5m)
- SimCLR ViT-T (11m)
- SimCLR ResNet18 (12m)
- SimCLR ViT-S (22m)
- SimCLR ResNet50 (26m)
- SimCLR ViT-B (88m)

- Supervised ImageNet ResNet50 (26m)
- SeCo ResNet50 (26m)
- DINO ImageNet ViT-S (22m)
- SimCLR ResNet50 (26m)
- SimCLR ViT-S (22m)

Little difference

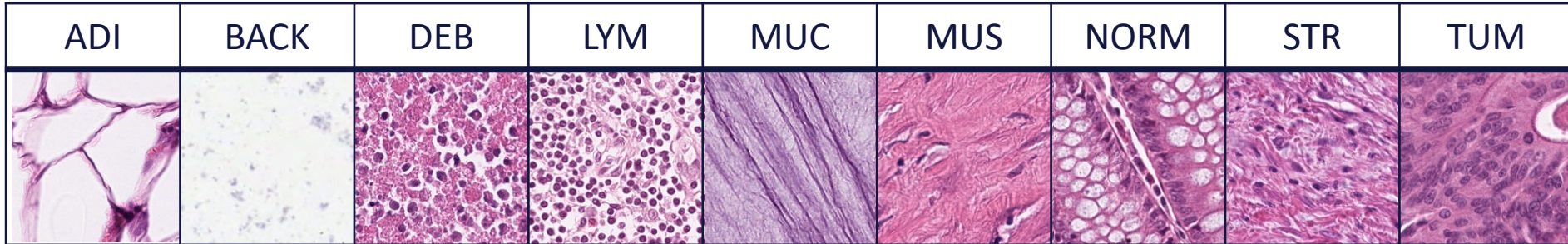
Smaller models sufficient for small training sets
Large models best for large training sets

Pre-training dataset matters a lot!

ImageNet: people, places, things
SeCo: satellite
EuroSAT: satellite

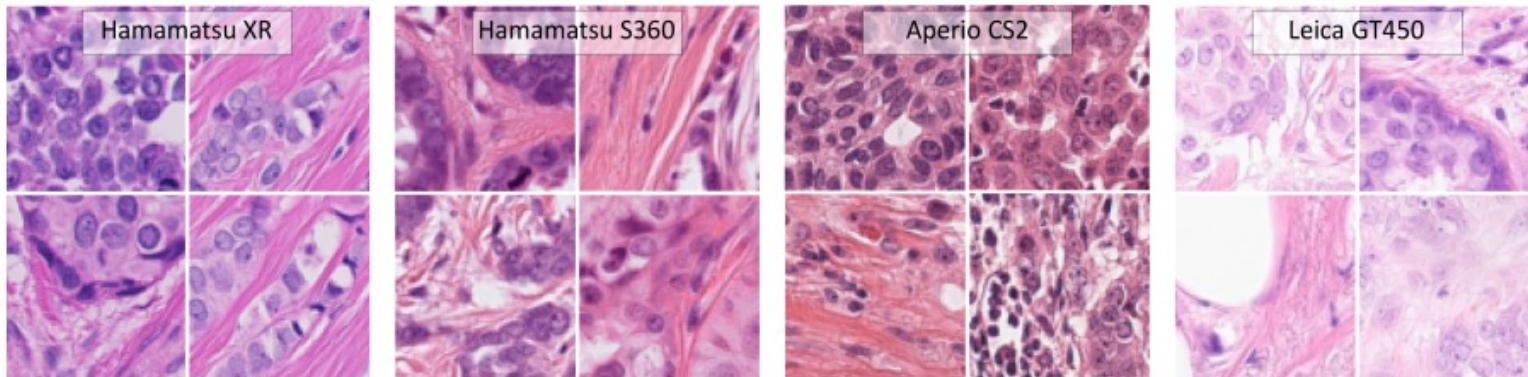
Example 2: Histopathology

- Domain: H&E colorectal tissue
- Training: 100k image patches
- Test: 7180 image patches from different hospitals
- Goal: predict 9 tissue classes
- Pre-train on various datasets, followed by linear classifier

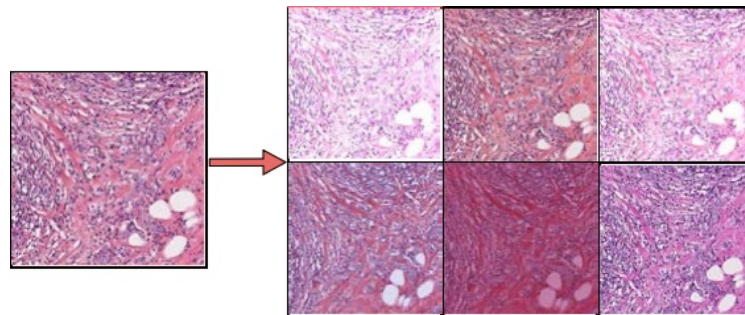


Example 2: Histopathology

Problem: color variations from different scanners or staining procedures

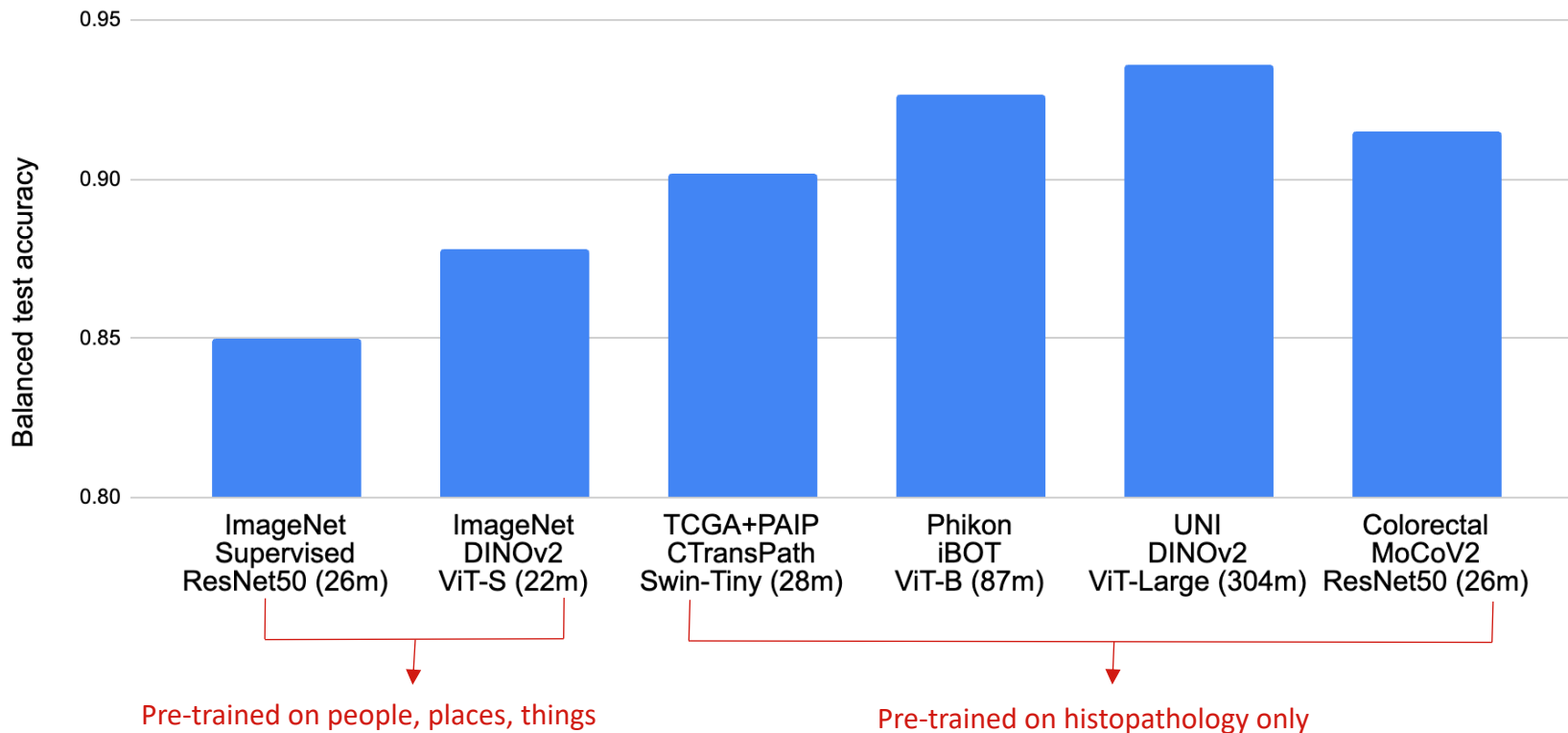


Solution: simulate color variations with image augmentation



Example 2: Histopathology

Domain-specific model improves generalizability



Pre-trained on people, places, things

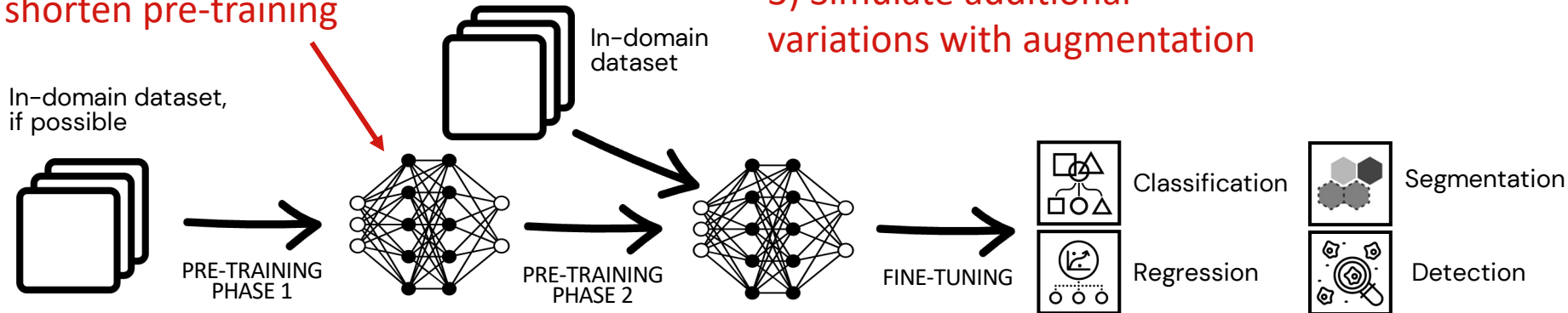
Pre-trained on histopathology only

Domain-Specific Foundation Model Best Practices

1) Start with another foundation model to shorten pre-training

2) Use a diverse dataset to capture variations within domain

3) Simulate additional variations with augmentation



Benefits of Domain-Specific Small Foundation Models

- 1) Domain-specificity allows for smaller models
- 2) Reduced computational needs for training and inference
- 3) Adaptable to multiple downstream tasks
- 4) Develop proof of concept quicker
- 5) Increased accuracy on downstream tasks
- 6) Less reliance on labeled data
- 7) Improved generalizability to distribution shifts

<https://pixelscientia.com/embedded2024/>

Links to these slides, articles, podcasts, and other resources to guide you on your journey.

Foundation Model ROI Workshop

Wednesday, June 5 @ 12 pm EDT/9 am PDT

A virtual workshop on how to identify the value and calculate the ROI of a vision foundation model approach.

Computer Vision Insights Newsletter

A biweekly newsletter that often features the latest research in foundation models.

Impact AI Podcast

Learn how to build a mission-driven machine, learning-powered company from the innovators and entrepreneurs who are leading the way.