

Automatic Image-Level Morphological Trait Annotation for Organismal Images

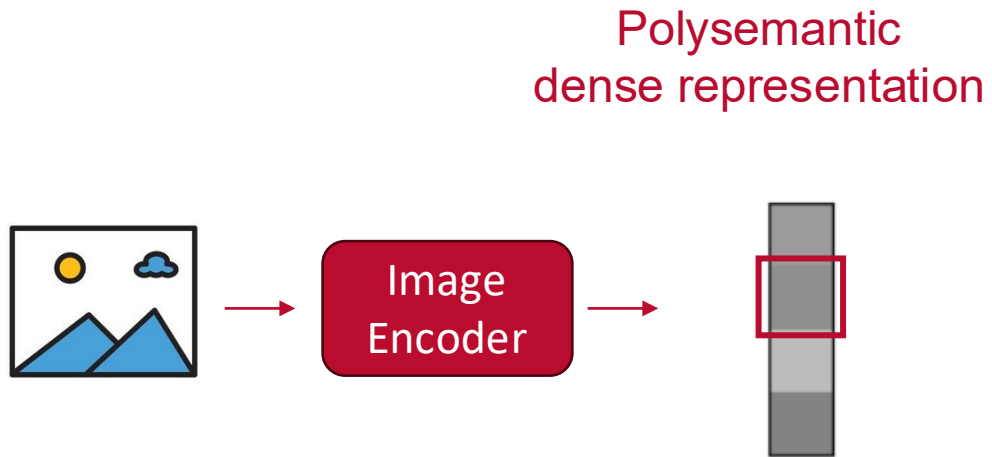
AUTOMATIC IMAGE-LEVEL MORPHOLOGICAL TRAIT ANNOTATION FOR ORGANISMAL IMAGES

Vardaan Pahuja¹ Samuel Stevens¹ Alyson East² Sydne Record² Yu Su¹

¹The Ohio State University ²University of Maine
pahuja.9@osu.edu



Discovering Latent Structure for Interpretable Visual Recognition



Discovering Latent Structure for Interpretable Visual Recognition

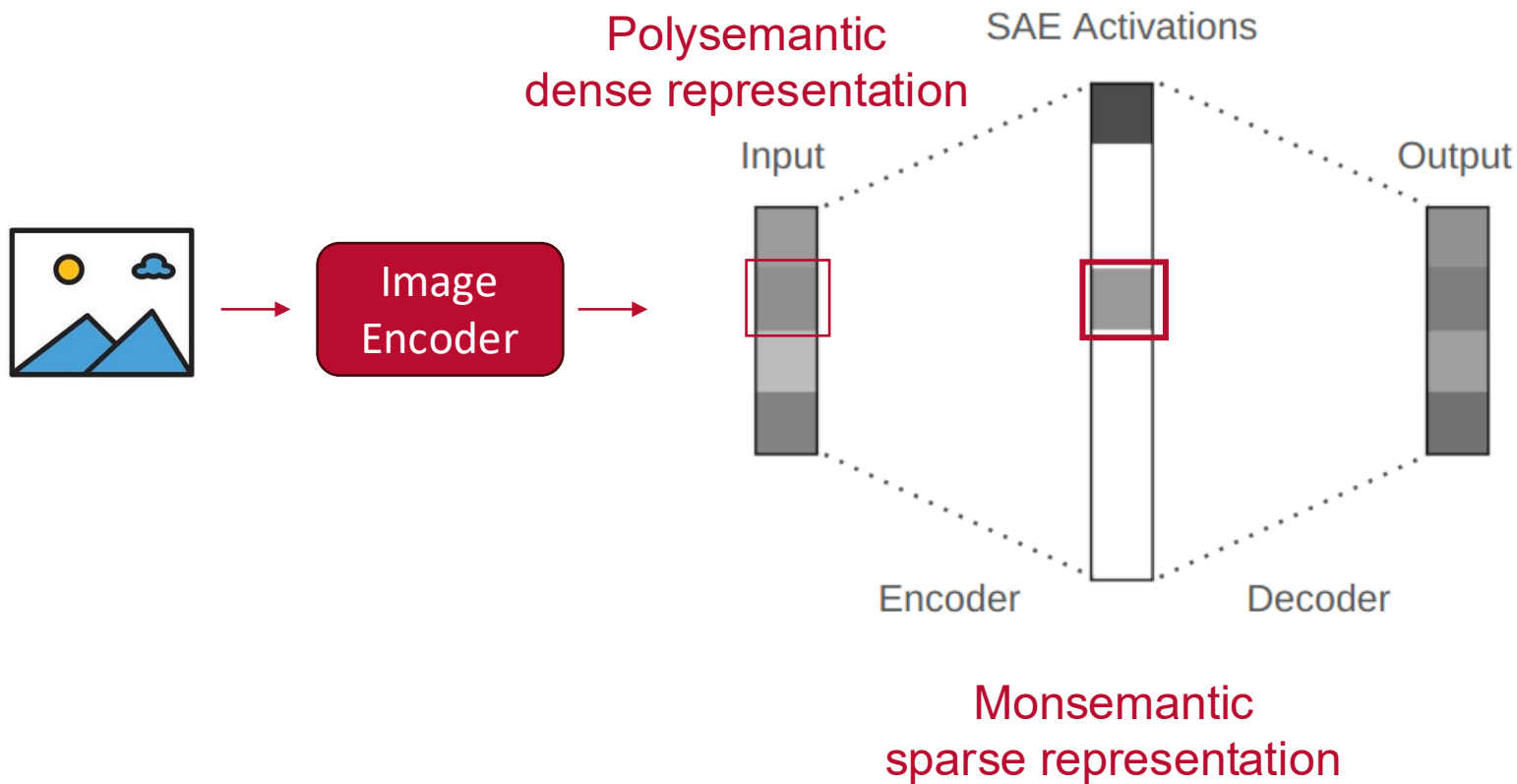
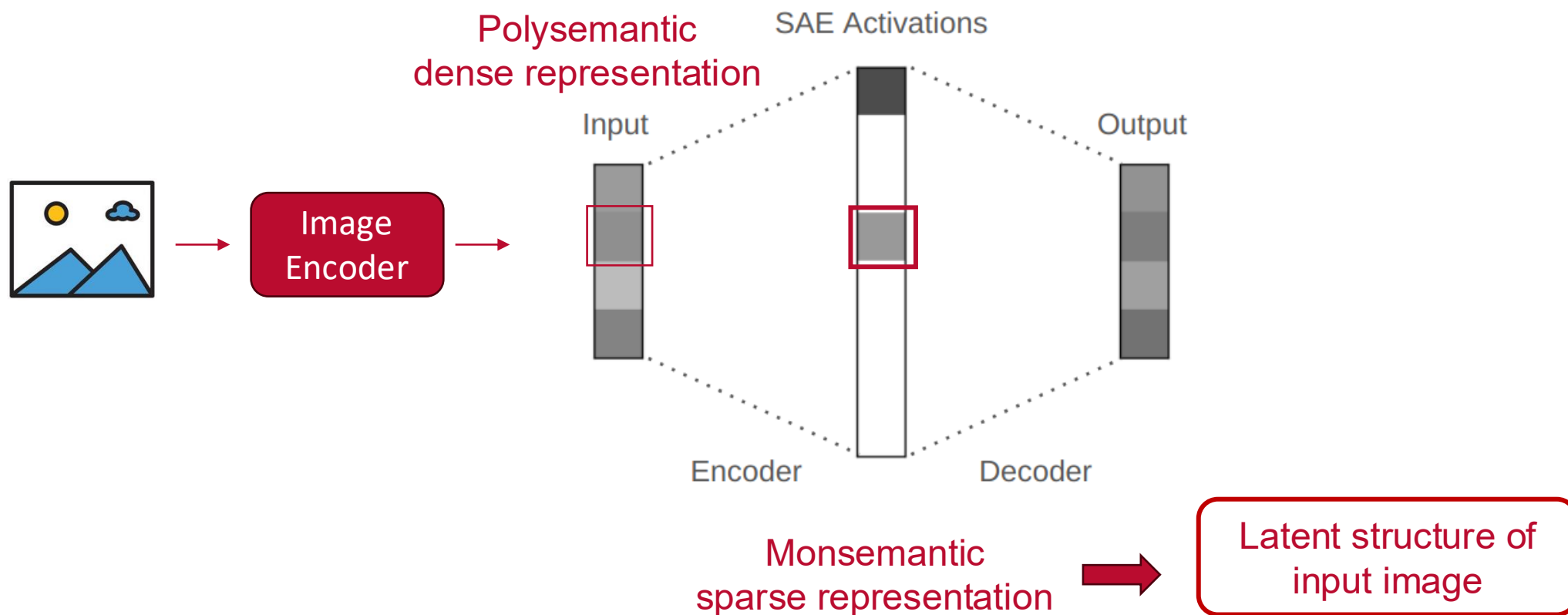


Fig. credit: Adam Karvonen



Discovering Latent Structure for Interpretable Visual Recognition



Discovering Latent Structure for Interpretable Visual Recognition

- Unlabeled data → dictionary of latent units
 - Sparsity
 - Non-Negativity
- Latent units whose activations map to spatially coherent regions

Learned SAE Latents Correspond to Biologically Plausible Traits

Neuron 4852



Cecidomyiidae



Stenomacrus



Cecidomyiidae



Stenomacrus



Bethylidae

Neuron 13860



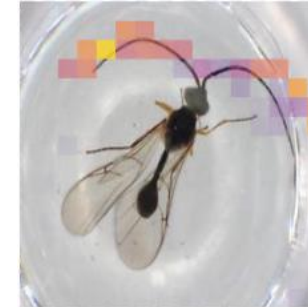
Scythrididae



Figitidae



Belytinae



Belytinae



Bethylidae



Learned SAE Latents Correspond to Biologically Plausible Traits

Neuron 4852



Cecidomyiidae



Stenomacrus



Cecidomyiidae



Stenomacrus



Bethylidae

Wings

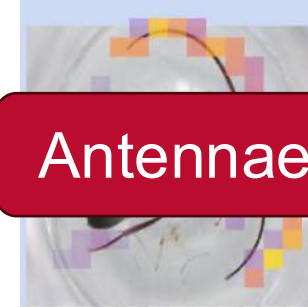
Neuron 13860



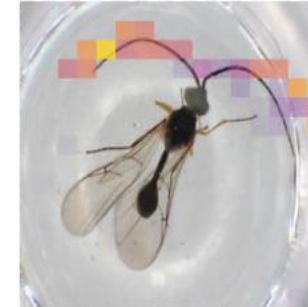
Scythrididae



Figitidae



Belytinae



Belytinae

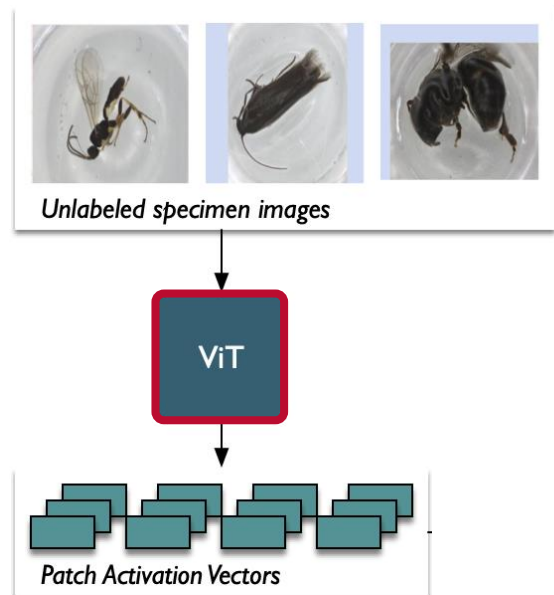


Bethylidae

Antennae

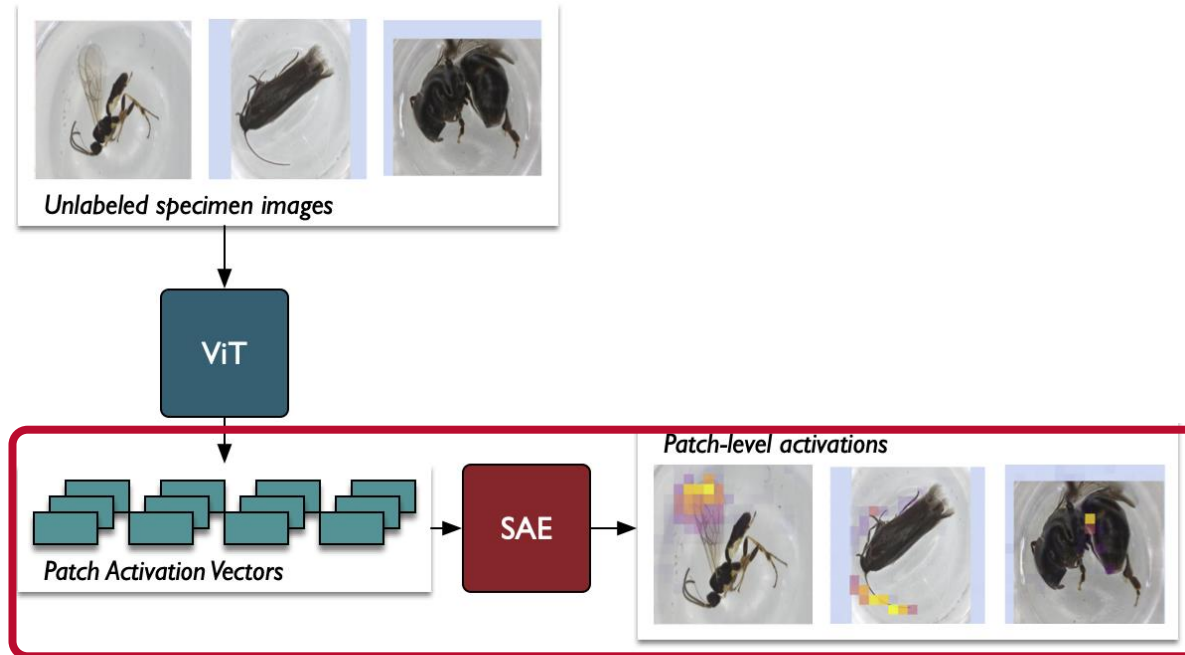


Morphological Trait Annotation Pipeline



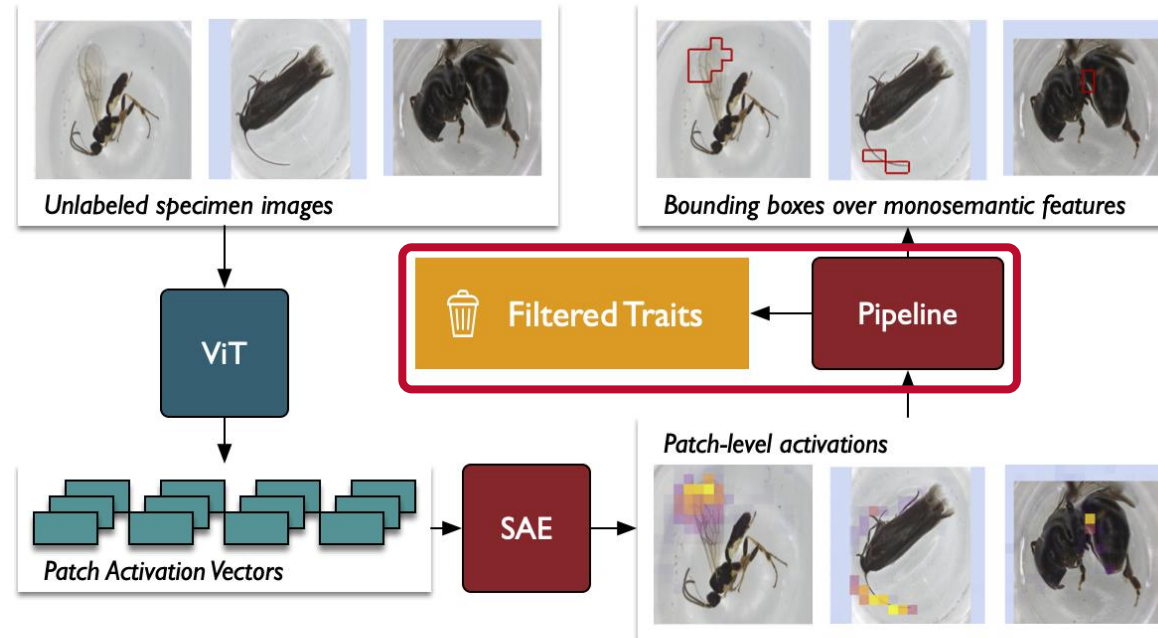
Compute dense
features using DINOv2

Morphological Trait Annotation Pipeline



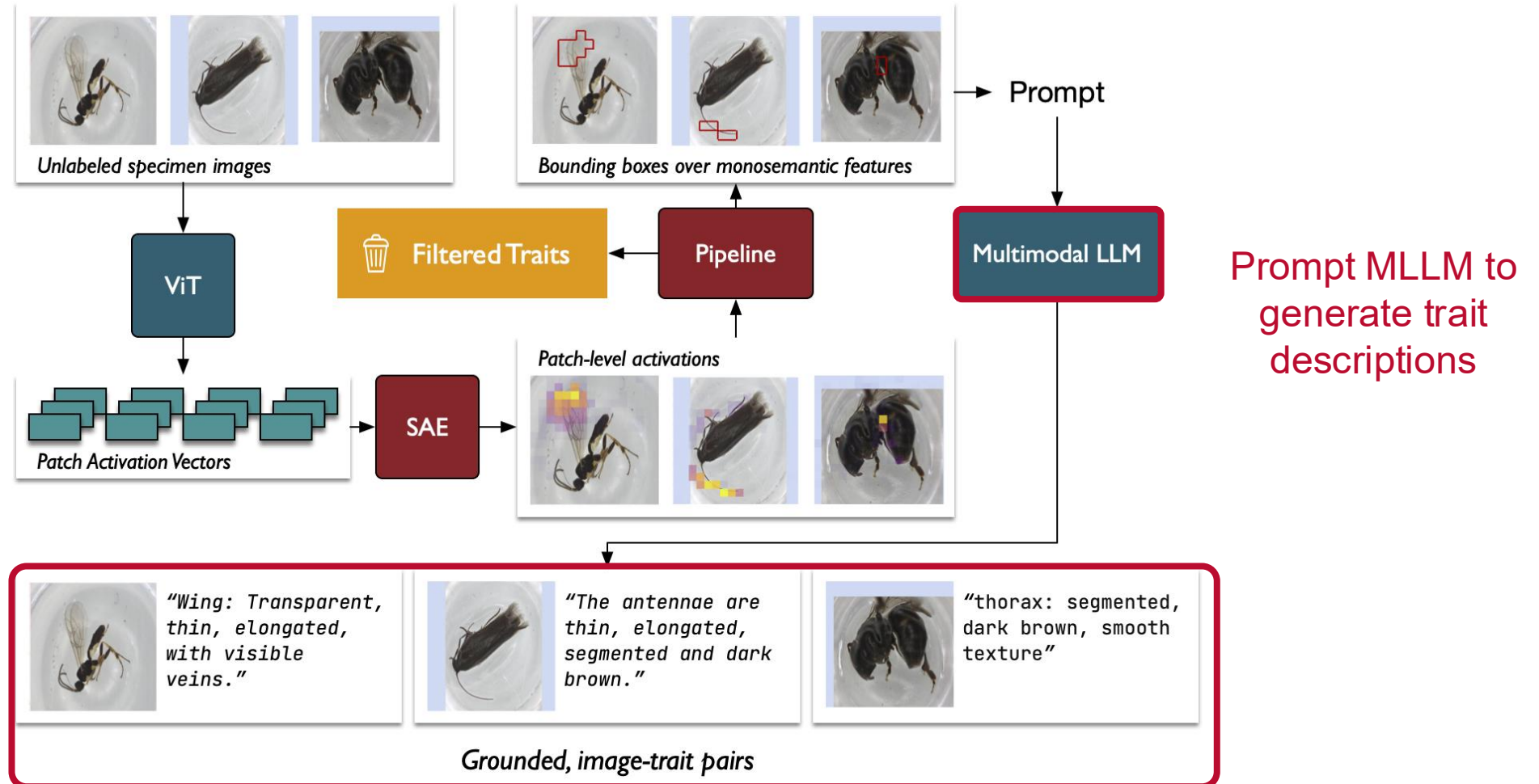
SAE identifies
high-activation latent units

Morphological Trait Annotation Pipeline



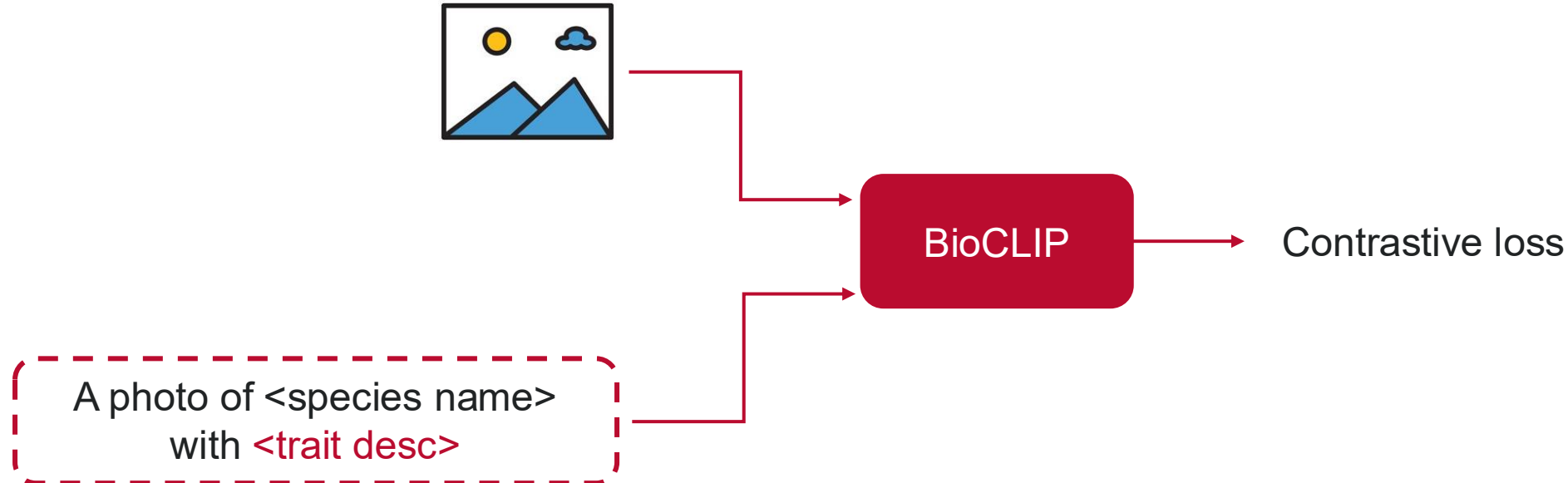
Species-contrastive
ranking: identify
species distinguishing
latents

Morphological Trait Annotation Pipeline



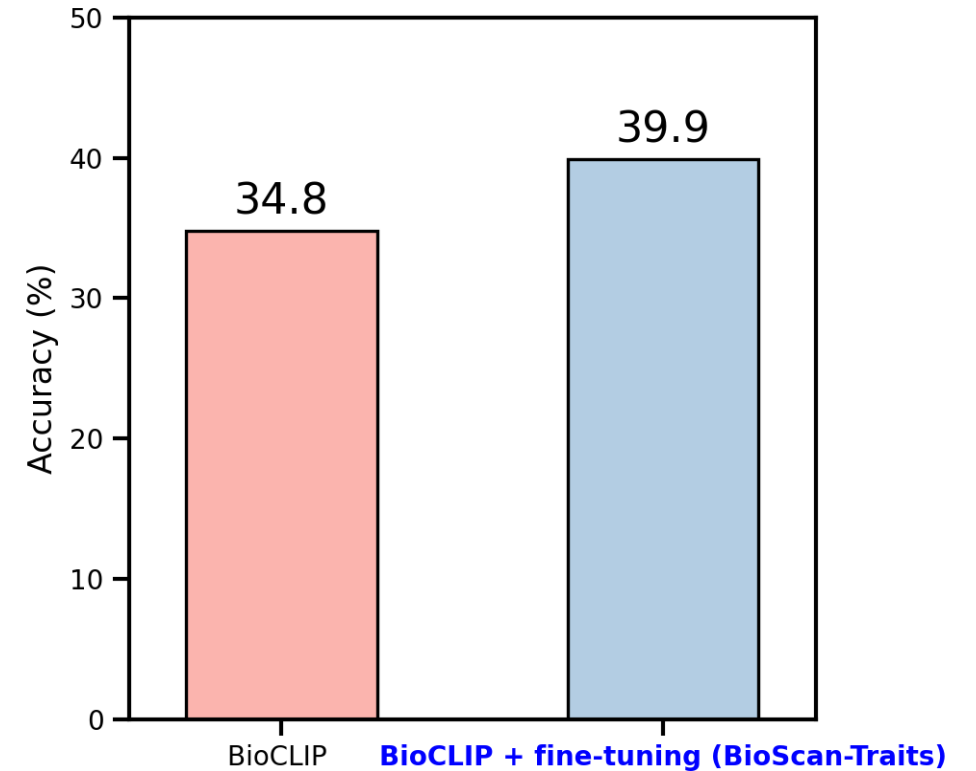
Trait-level supervision improves OOD generalization

- **BioScan-Traits**: 80K morphological trait annotations across 19K insect images from BioSCAN-5M.
- Finetune BioCLIP on the dataset.



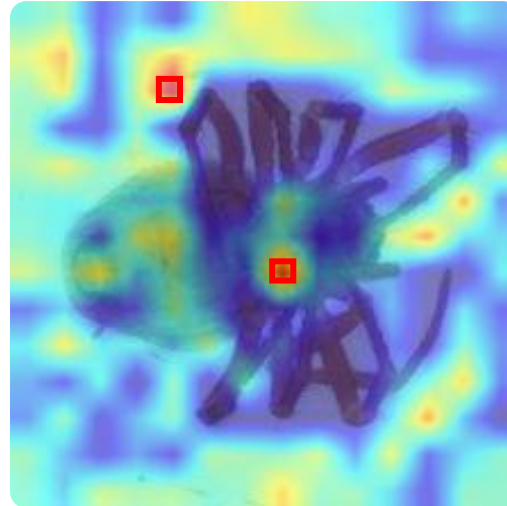
Trait-level supervision improves OOD generalization

- **BioScan-Traits**: 80K morphological trait annotations across 19K insect images from BioSCAN-5M.
- Finetune BioCLIP on the dataset.
- Evaluation on Insects, a species classification benchmark.



Comparison with GradCam

Grad-Cam



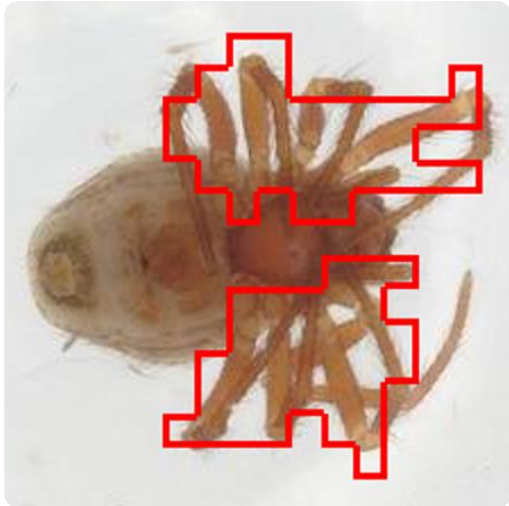
Trait Descriptions

Grad-Cam: **Head:** Highlights the cephalothorax, emphasizing its rounded shape and the presence of sensory organs. **Legs:** Focuses on the anterior legs, showcasing their robust structure and segmentation. - **Body:** Emphasizes the abdomen, noting its segmented appearance and the presence of spinnerets at the posterior end.

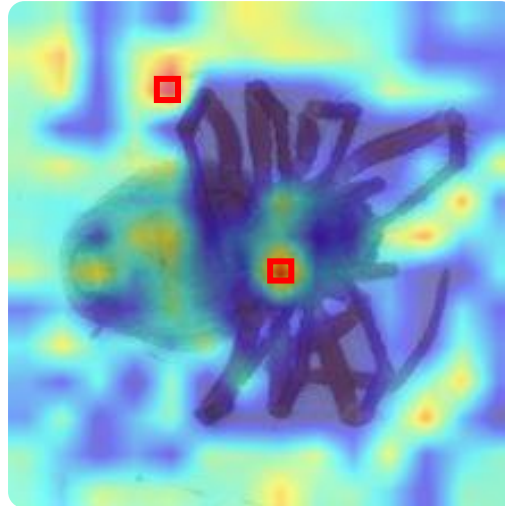
- ✗ Interpretable
- ✗ Species-discriminative

Comparison with GradCam

BioScan-Traits



Grad-Cam



Trait Descriptions

Grad-Cam: **Head:** Highlights the cephalothorax, emphasizing its rounded shape and the presence of sensory organs. **Legs:** Focuses on the anterior legs, showcasing their robust structure and segmentation. - **Body:** Emphasizes the abdomen, noting its segmented appearance and the presence of spinnerets at the posterior end.

Bioscan-Traits: **Leg:** Thin, elongated, light brown, segmented.

- ✓ Interpretable
- ✓ Species-discriminative

Key Takeaways

- **SAEs as part/trait detectors**: Sparse activations yield localized, monosemantic features aligned with morphological traits.
- **Latents → traits**: Select salient latents and verbalize them into natural-language trait descriptions.
- **Why it matters**: Trait supervision improves generalization on a downstream species classification benchmark.