InSpect

A Curated Natural History Collection Dataset for Insect Specimen Understanding

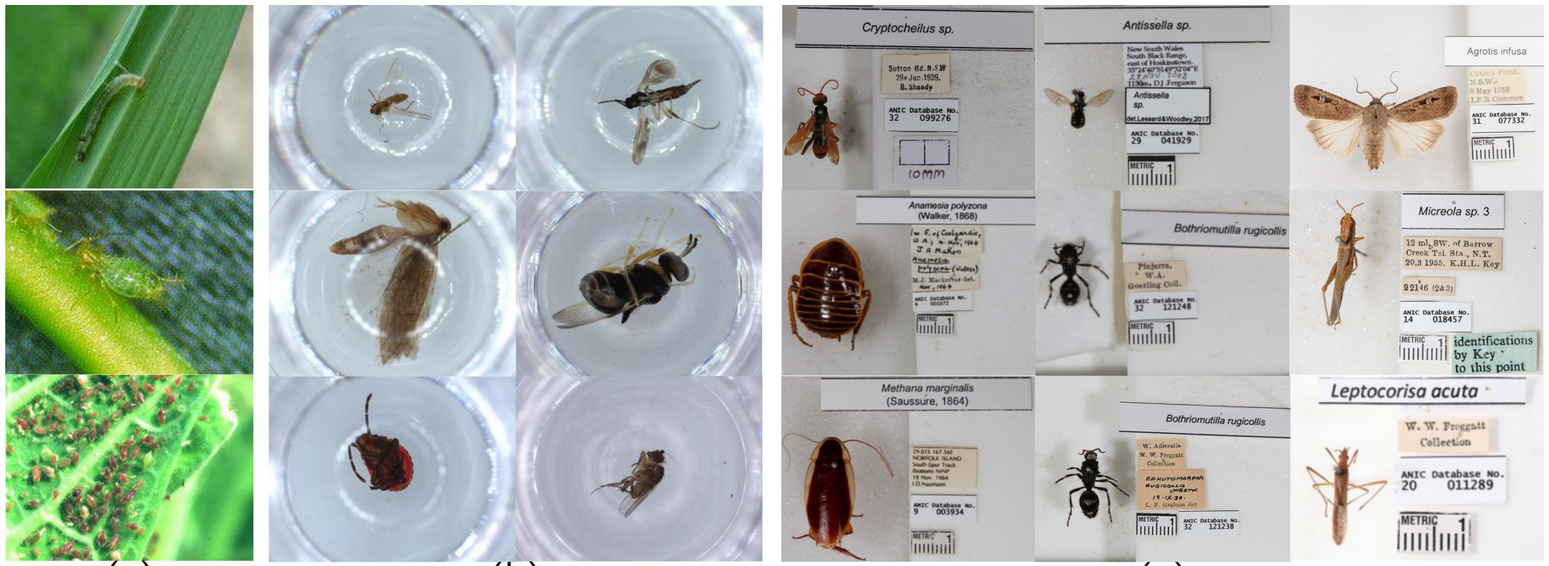
Evaluations & Datasets Track Spotlight



Junhao Zhang

Technology, CSIRO
Imaging and Computer Vision Group

Why InSpect?



(a) (b) (c)

(a) IP102 (b) BIOSCAN-5M (c) InSpect

Taxonomic recognition

Recognize taxa beyond the training set

Anatomical localization

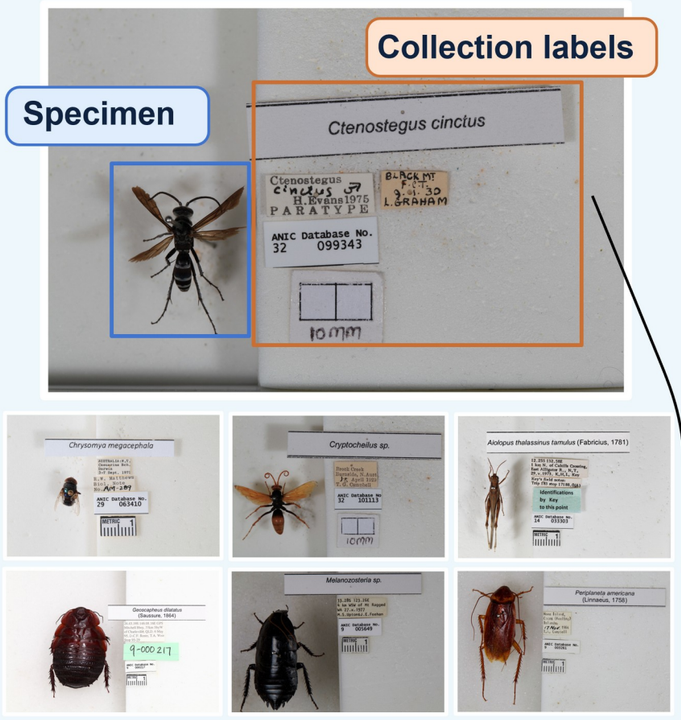
Separate antennae, legs, wings and body regions

Collection photographs preserve specimens and their physical labels.

Aligned collection data and annotations

(a) Collection images

Museum specimens with collection labels.



Illustrative specimen thumbnails

48,184	653	1,859
images	taxa	part-annotated images

(b) Aligned data and annotations

Each specimen is linked with taxonomy and metadata, a subset is with part masks.

Hierarchical taxonomy

Order	Hymenoptera
Family	Pompilidae
Genus	Ctenostegus
Species	Cinctus

Structured metadata

Locality	Black Mt
Collection date	9 Jan 1930
Collector	L. Graham
Sex	Male
Determiner	H. Evans
Determination date	1975
Notes	PARATYPE

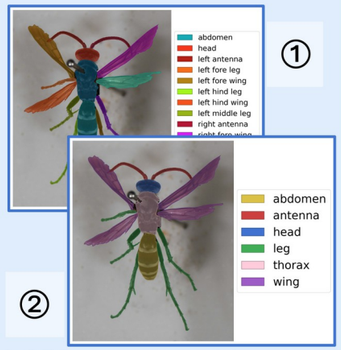
From collection labels

From specimen image



Part annotation

Anatomical part masks



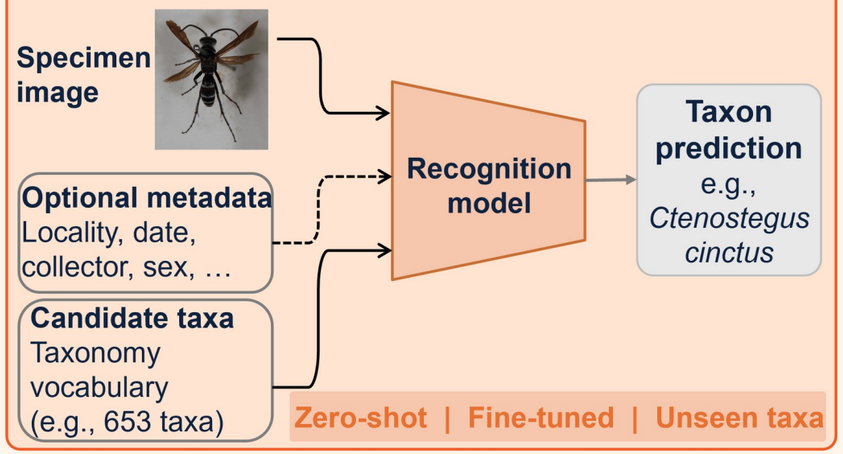
Specimen crop

1,859 annotated images

(c) Two benchmarks

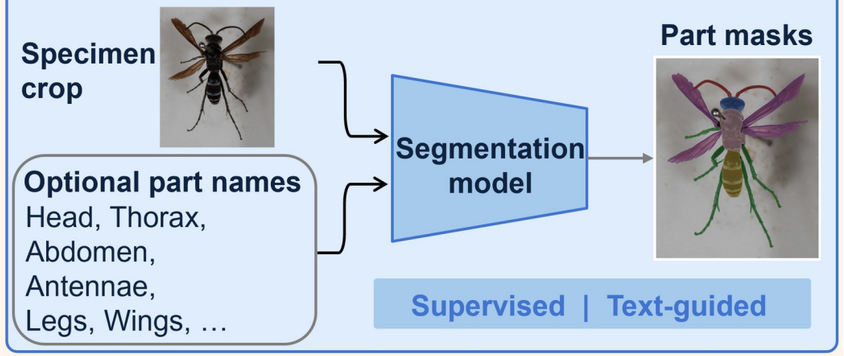
Benchmark I: Open Taxonomic Recognition

Recognize the specimen's taxon from image and metadata.



Benchmark II: Anatomical Part Segmentation

Predict anatomical part masks for a specimen crop.



Benchmark I: open taxonomic recognition



Specimen crop

Rank candidate taxon names from the specimen crop.

Evaluation	Candidate taxa	Relationship to training
Test-seen	517	Taxa observed in training
Test-unseen	136	Taxa held out from training

Seen images: ~70% train / 10% val / 20% test-seen

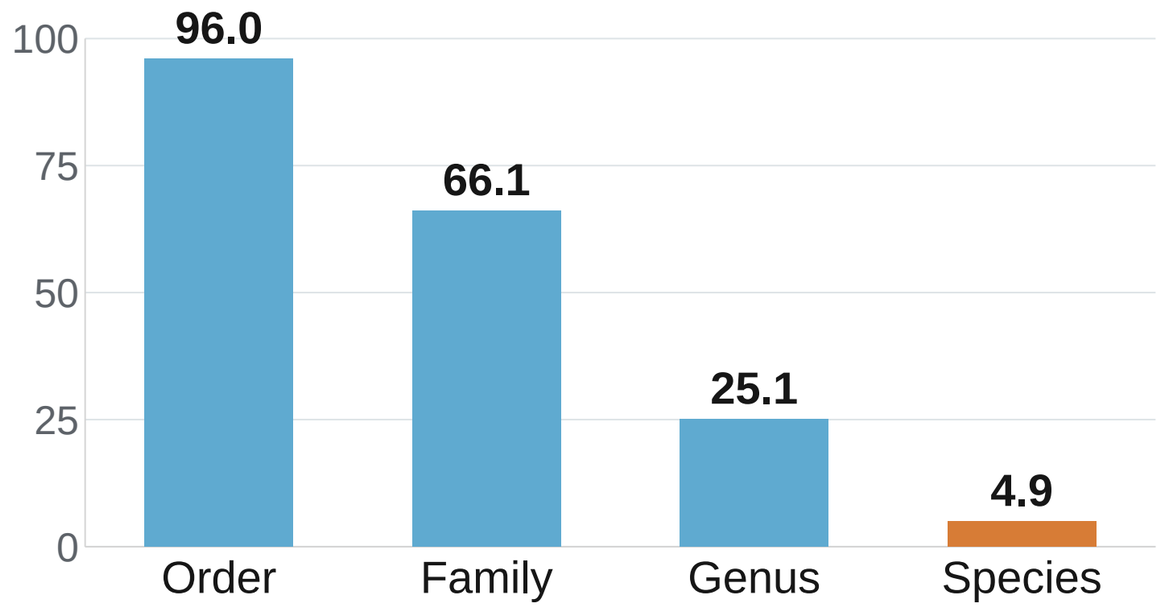
27,082 train 3,867 val 7,759 test-seen
9,476 images in test-unseen

Separate candidate pools mean seen/unseen scores are not a controlled difficulty comparison.

Fine taxonomic recognition remains difficult

Zero-shot hierarchy

SigLIP 2, test-unseen accuracy (%)



Adaptation on unseen taxa

SigLIP 2 setting	Top-1	Top-5
Zero-shot	4.93	14.86
Image fine-tuning	6.16	21.21
Image + metadata fine-tuning	11.07	32.32

Species accuracy (%), Table 4.
Metadata assists training only. Inference uses images.

Fine taxa remain challenging, even when context helps.

Benchmark II: anatomical part segmentation

Specimen crop



Singular labels



Spatial-aware labels



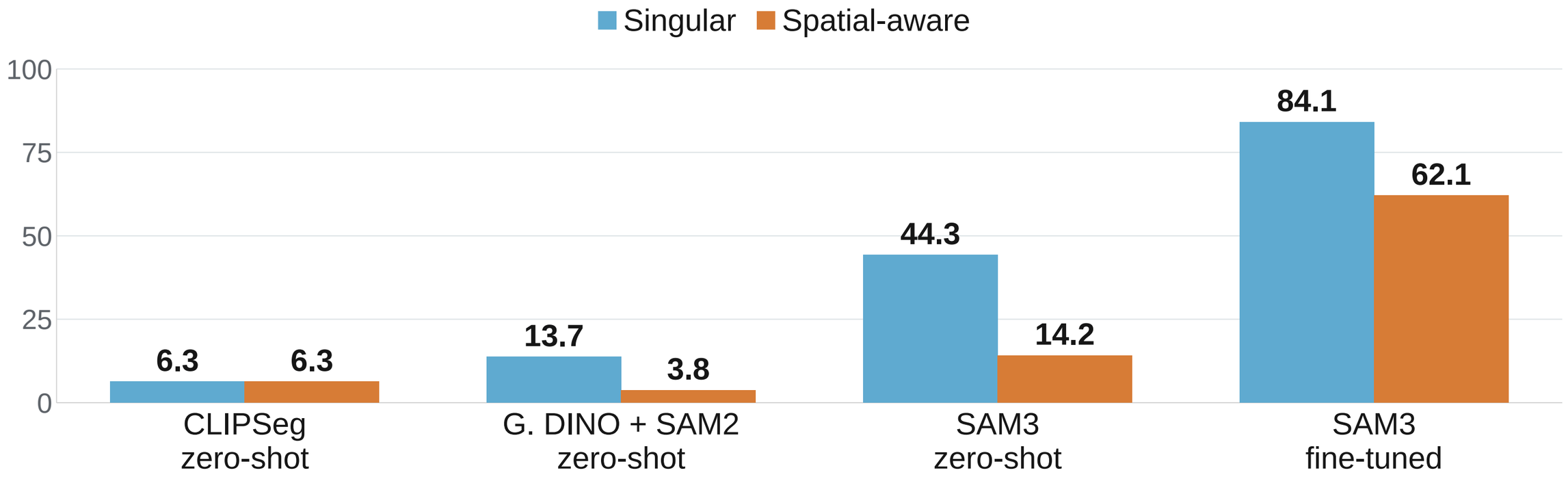
11 labels: leg, antenna, wing, ...

31 labels: left/right, fore/middle/hind, ...

1,859 annotated images covering 632 taxa

Text-guided segmentation benefits from adaptation

Text-guided part prompts. Metric: mIoU (%)



Fine-tuning substantially improves anatomical part localization.

Table 6. Label settings use different image splits, so the cross-setting gap is not a controlled granularity comparison.

Qualitative part segmentation

Crop	Ground truth	SAM3 zero-shot	SAM3 fine-tuned
			

abdomen antenna head leg thorax wing

Zero-shot misses or merges parts. Fine-tuning improves localization.

InSpect

Collection context and specimen morphology for recognition and anatomical understanding



Paper, dataset and code

csiro-sd.github.io/InSpect

Thank you