

Exploring Vision Foundation Models for Analysis of Anatomical Micro-CT Volumes through Sample Reorientation

S.K. Mackie, MF.F. Petersen, W.M. Laprade, and H.M. Kjer

Background and methodology

Micro-CT scanning enables the creation of high-fidelity 3D digital twins of natural heritage collections. Many types of analysis on these complex 3D volumes remains challenging to do and often depend mostly on information from the specimen surface.

We propose a rendering-based workflow that transforms a 3D problem into a series of 2D tasks. By generating multiple rendered views of a specimen, we can leverage vision foundation models (DINO) for feature extraction. Resulting 2D matches are subsequently triangulated to recover information in 3D space.

We demonstrate the approach on micro-CT scans of bugs and tackle the problem of specimen orientation estimation.

Dataset and experiment

The BugNIST dataset contains micro-CT scans of individual insects from 12 taxa of varied anatomical complexity and unknown poses. The objective is to estimate the 3D head-to-centroid axis, enabling consistent reorientation of specimens into a common reference frame (e.g., with the head directed upward). Performance is assessed through qualitative evaluation on a test set comprising 10 specimens from each class and compared against a classic 3D PCA-based realignment.

We investigate the influence of:

- Segmentation and preprocessing strategies of the input micro-CT volume.
- Foundation model size and rendering parameters, including image resolution and number of rendered views.
- Strategies for 3D triangulation to achieve robust matching.

Key finding

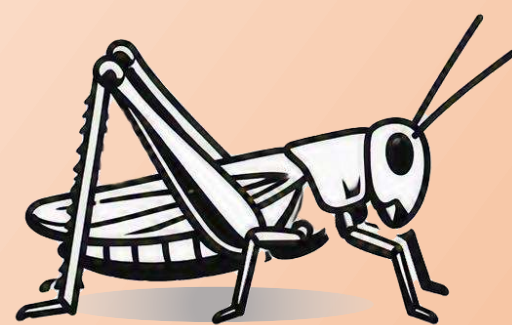
The proposed approach works well for feature-rich (presence of wings, leg etc.) specimens, but struggle with simple anatomies (cylinder-like larva and pupa).



UNIVERSITY OF COPENHAGEN



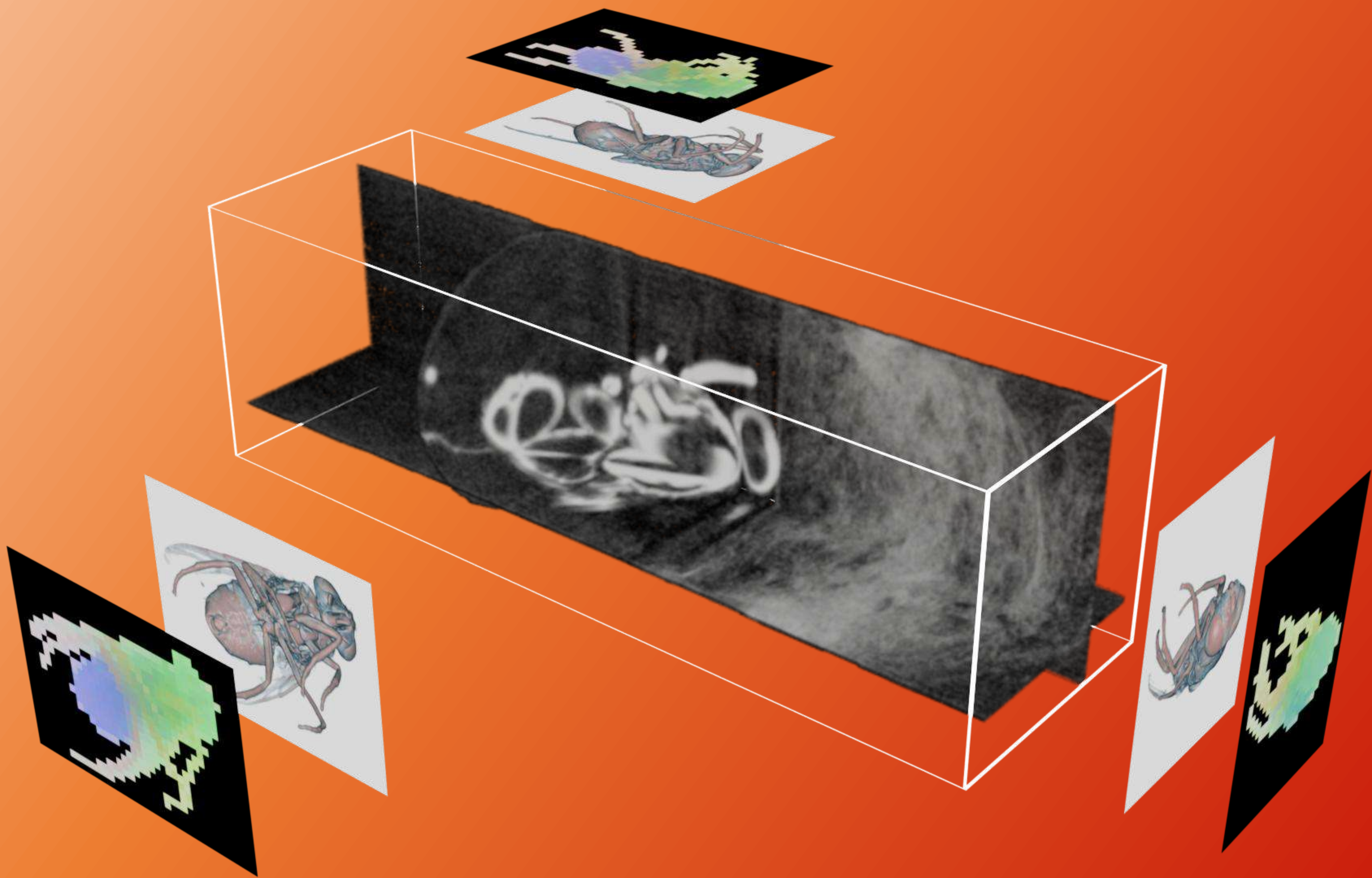
LUND UNIVERSITY



3D Bugs + 2D DINO:



A solution to 3D **specimen orientation** in **micro-CT** scans of **natural heritage** samples (Bugs) utilizing 2D vision foundations models (DINO)



The paper



Our method step-by-step



Figure 1 – Overview: 2D rendering of raw (1) and segmented CT (2). Detection of head (3) across multiple views to estimate 3D main axis, enabling specimen reorientation (4).

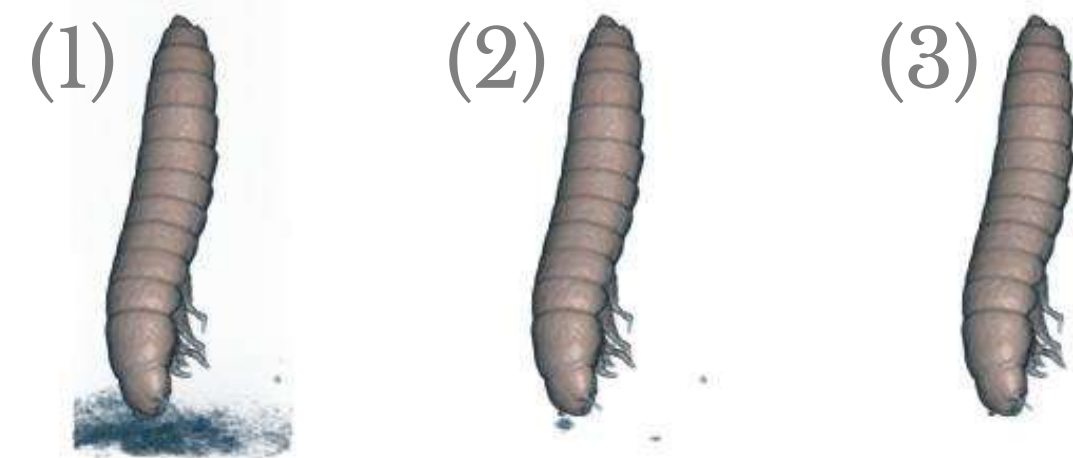


Figure 2 – Step 1: Raw rendering (1) of meal worm CT, Otsu threshold (2) and largest component (3).

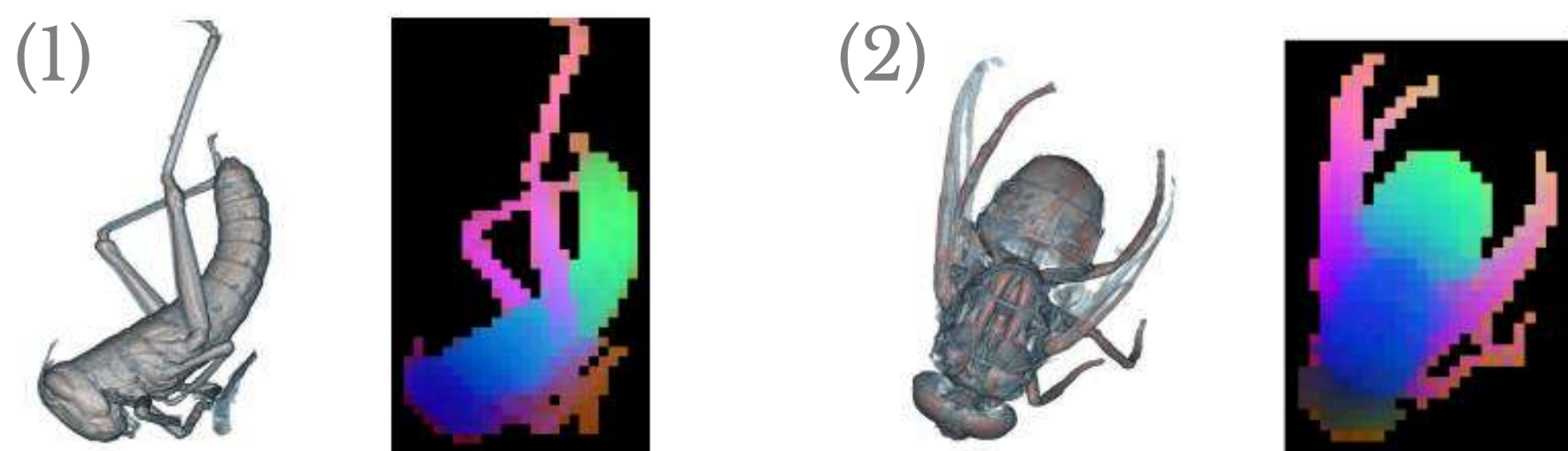


Figure 3 – Step 2: Shared PCA-decomposition of DINO features from renders of a grasshopper (1) and a blow fly (2).

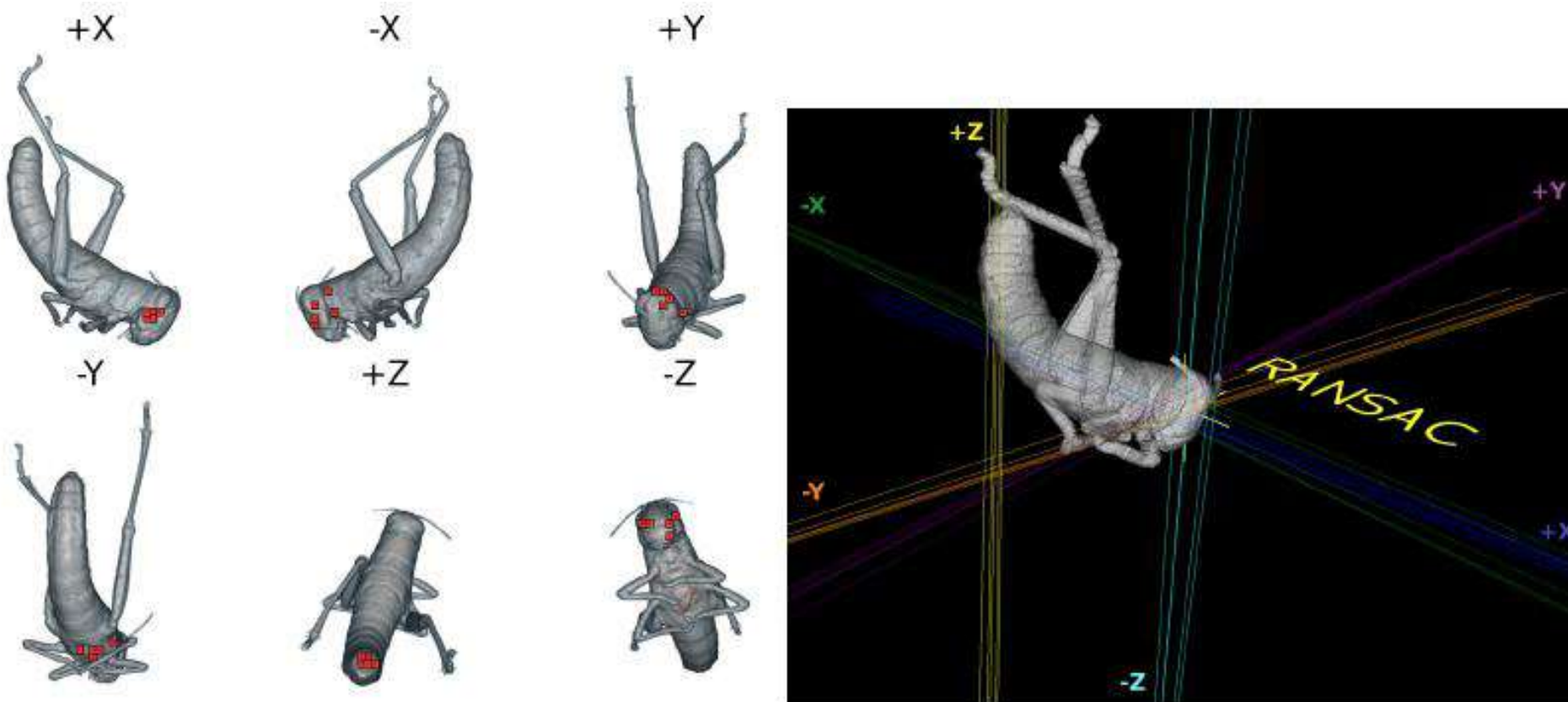


Figure 4 – Step 3: Head detections on 6 rendered views (left). Least squares triangulation of head position in 3D space using RANSAC (right).

Experimental results:

Baseline model: Otsu segmentation, 6 axis-aligned views, the Big DINO model (ViT-B/16) in 60x60 patches, matching the 5 most similar patches against an average head-feature vector from 5 “training samples” of each species.

Configuration	AC	BC	BF	BL	BP	CF	GH	MA	ML	PP	SL	WO	Sum	Accuracy
PCA	10	9	0	6	7	2	2	8	4	3	5	5	61	51%
Baseline	10	10	10	5	10	10	10	2	8	3	5	4	87	73%
1 image annot.	10	9	6	5	7	8	8	2	4	3	5	1	68	57%
1 animal annot.	10	10	10	5	7	8	5	2	4	3	5	2	71	59%
Model ViT-S/16	10	10	9	5	10	8	9	2	10	2	5	4	84	70%
Model ViT-L/16	10	10	10	5	10	10	10	2	9	3	5	10	94	78%
Model ViT-H+/16	10	10	10	10	10	10	10	2	10	3	5	10	100	83%
Grid 30	10	10	10	5	10	10	10	2	7	3	5	4	86	72%
Grid 90	10	10	10	5	10	10	10	2	6	3	5	3	84	70%
Threshold p_{97}	5	6	1	5	10	6	6	2	7	3	5	3	59	49%
Threshold p_{98}	6	8	3	5	7	3	9	2	8	3	5	3	62	52%
Threshold $p_{99.5}$	10	10	10	4	10	10	10	2	4	3	4	3	80	67%
Top-k 3	10	10	10	5	10	10	10	2	8	3	5	3	86	72%
Top-k 8	10	10	10	5	10	10	10	2	8	3	5	3	86	72%
Four views (pyramid)	10	10	10	5	7	10	10	2	4	3	5	3	79	66%

Table 1: Alignment success rate (10 test samples per specie) for different experiment conditions.

AC=brown cricket, BC=black cricket, BF=blow fly, BL=Buffalo larva, BP=blow fly pupa, CF=curly-wing fly, GH=grasshopper, MA=maggot, ML=mealworm, PP=bottle fly pupa, SL=soldier fly larva, WO=woodlice