

DTU



Computer Vision for Natural Heritage workshop, ECCV 2026



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

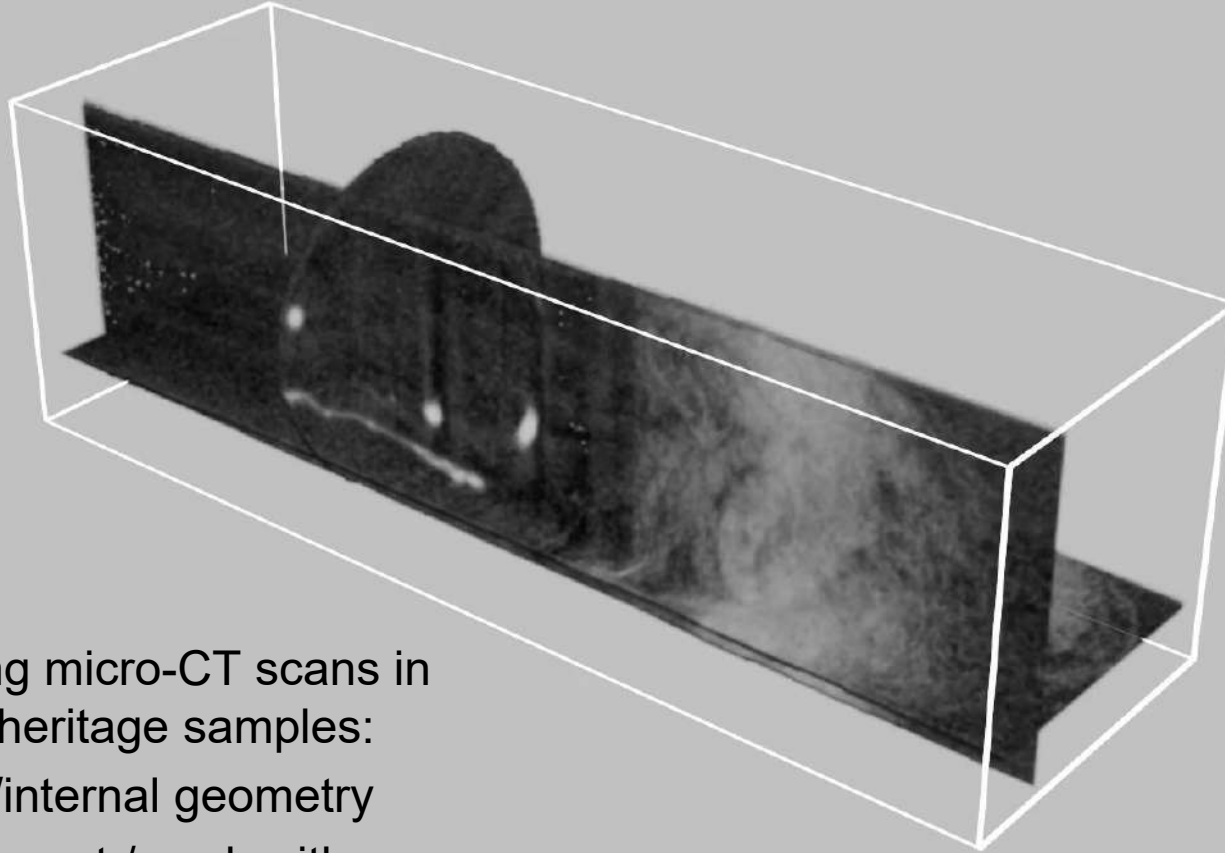
Vision Foundation models



- **Vision foundation models** (e.g. the line of DINO models)
 - Informally: Possess strong visual intelligence or semantic understanding,
 - Technically: Can extract highly useful features from 2D images, allowing for “solutions” to a variety of problems in a few- or zero-shot manner,
 - Still just for 2D (natural) images

M. Oquab et al, "DINOv2: Learning Robust Visual Features without Supervision", 2024

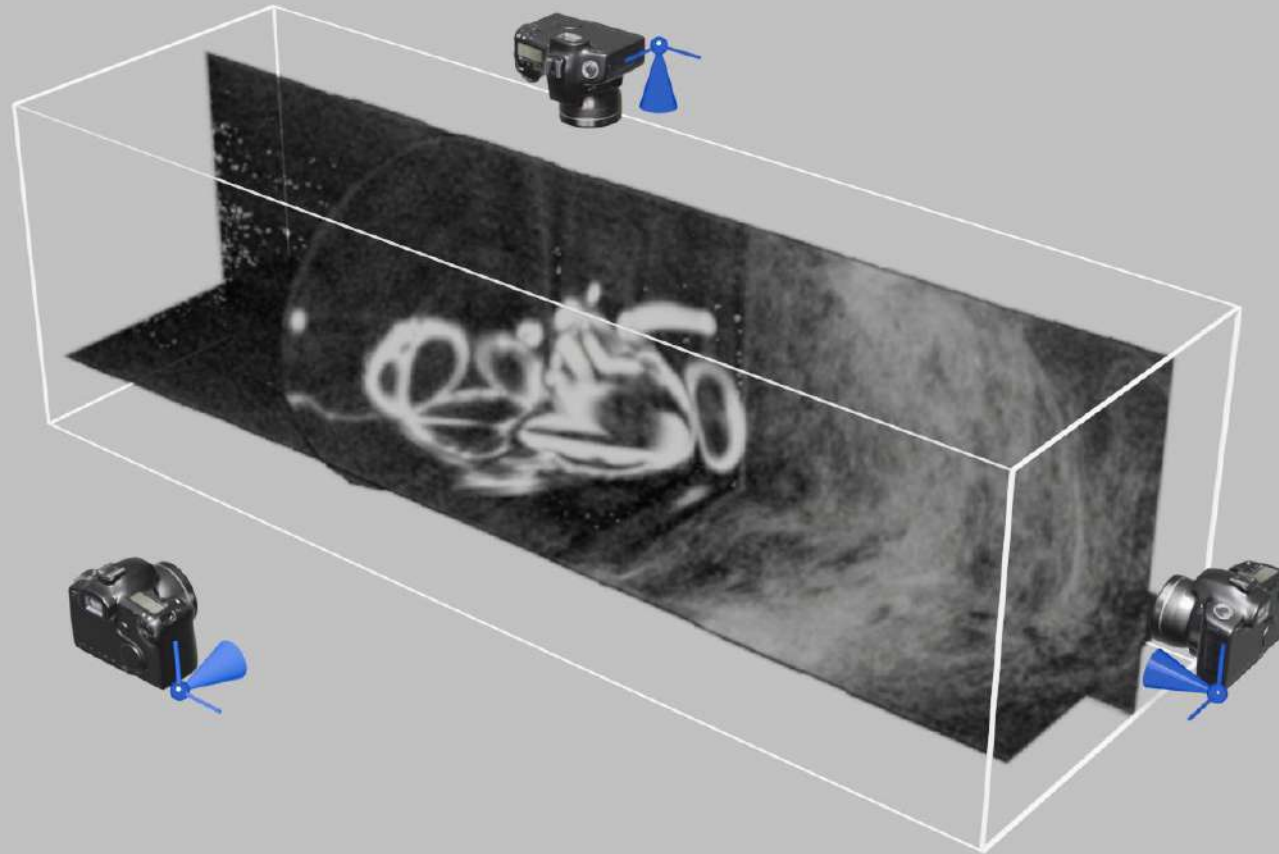
Natural heritage and micro-CT



- There is a case for using micro-CT scans in digitalization of natural heritage samples:
 - + It is 3D > full interior/internal geometry
 - % It is 3D > hard to interpret / work with
- AND we do not have 3D foundation models

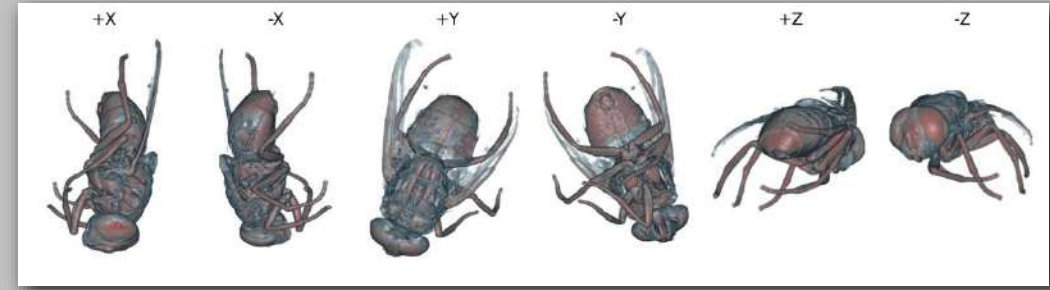
- Would you be able to see what the scanned object is?

1. Rendering: (3D > 2D)



- Use **computer graphics** to turn 3D analysis into several 2D ones?
- Place N cameras with known properties and position

1. Rendering: (3D > 2D)

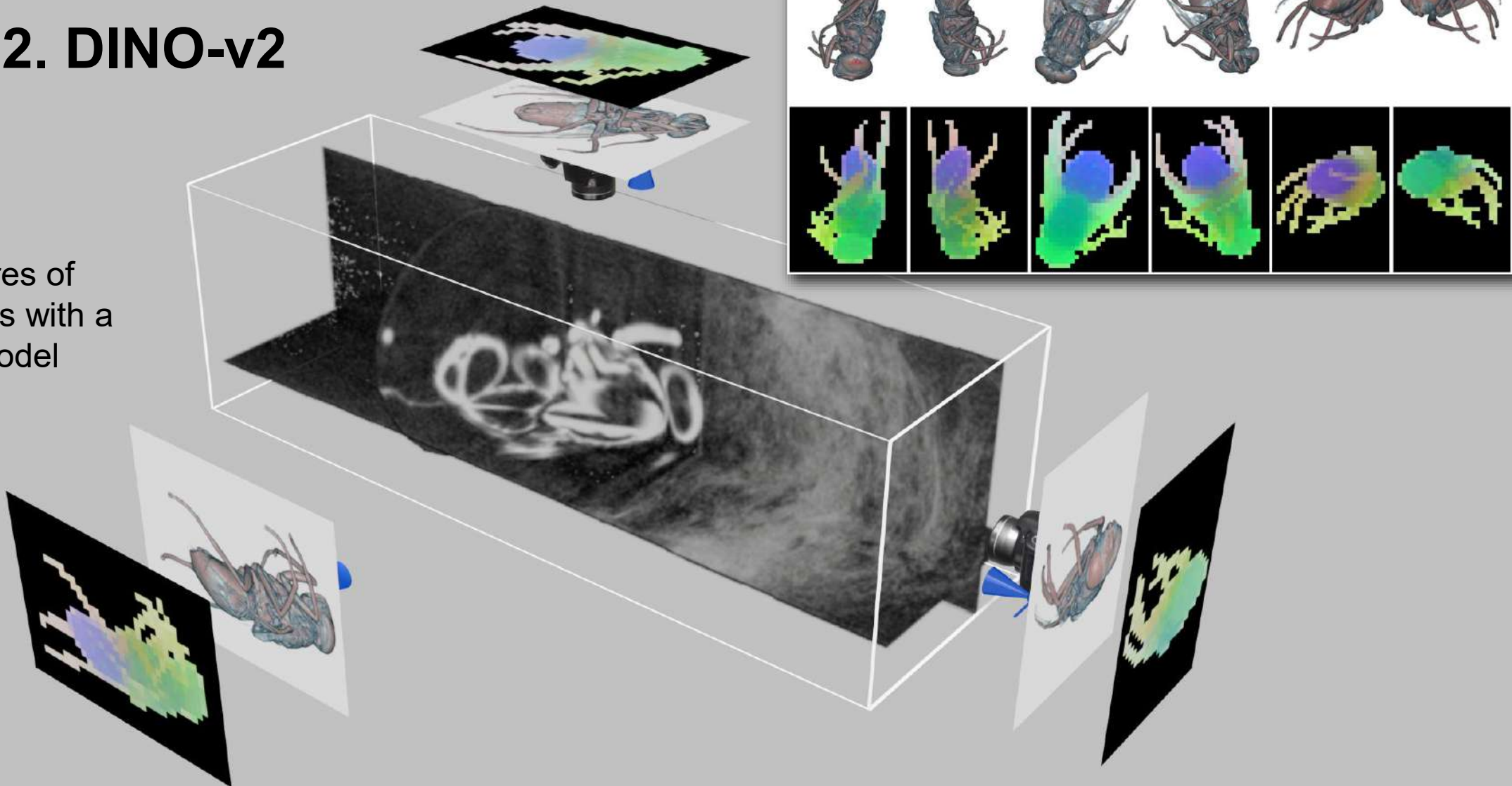


- Use **computer graphics** to turn 3D analysis into several 2D ones?
- Place N cameras with known properties and position.

- Would you be able to see what the scanned object is?

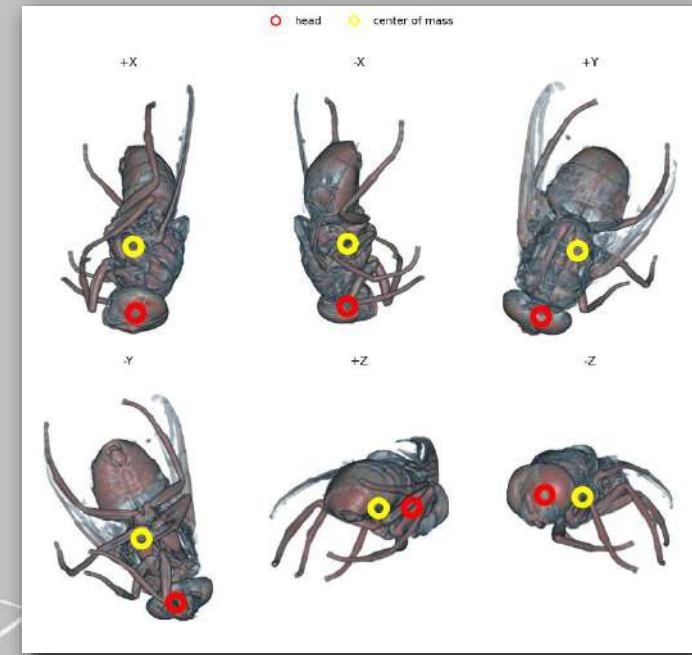
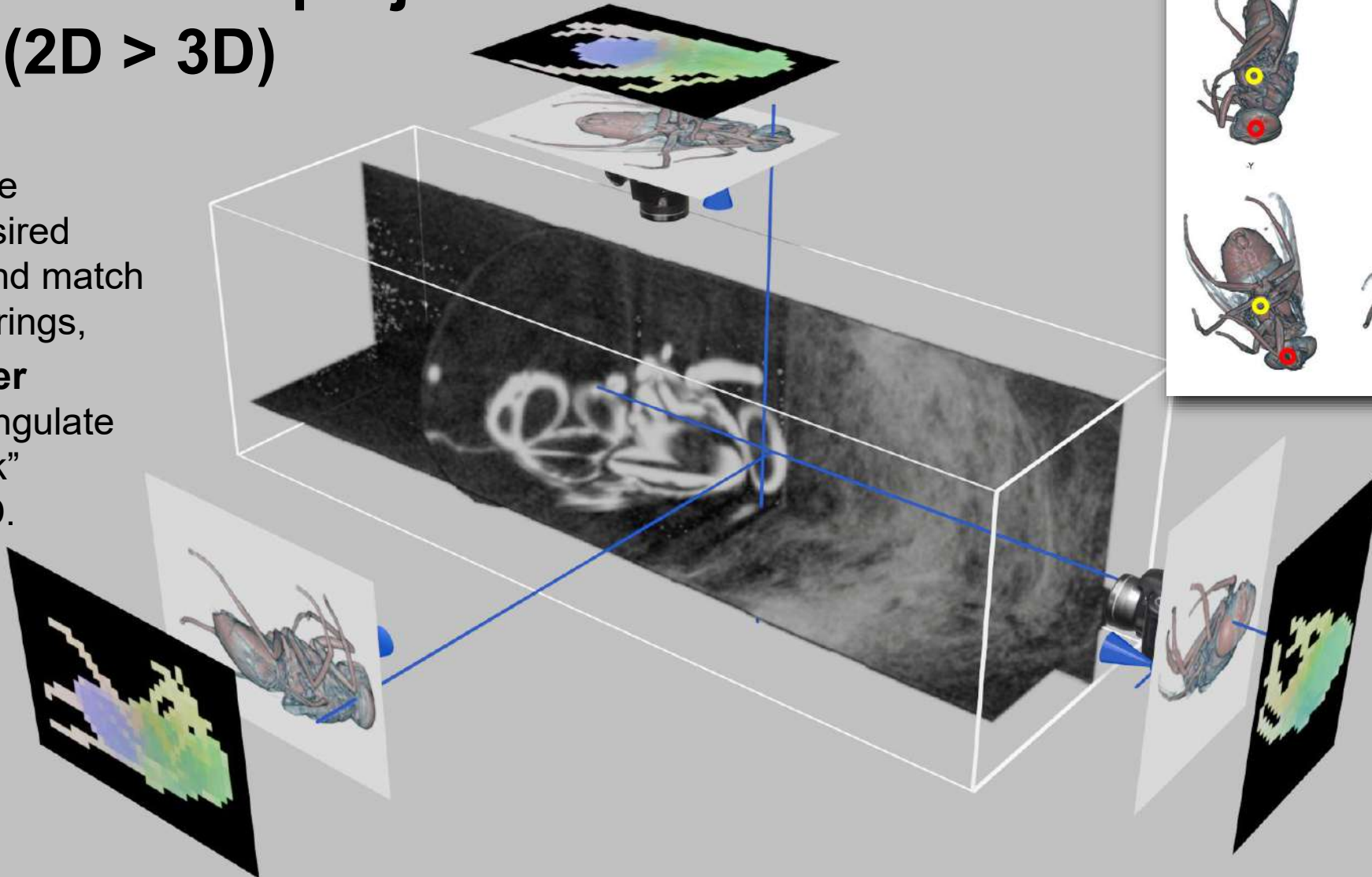
2. DINO-v2

- Extract features of the renderings with a foundation model



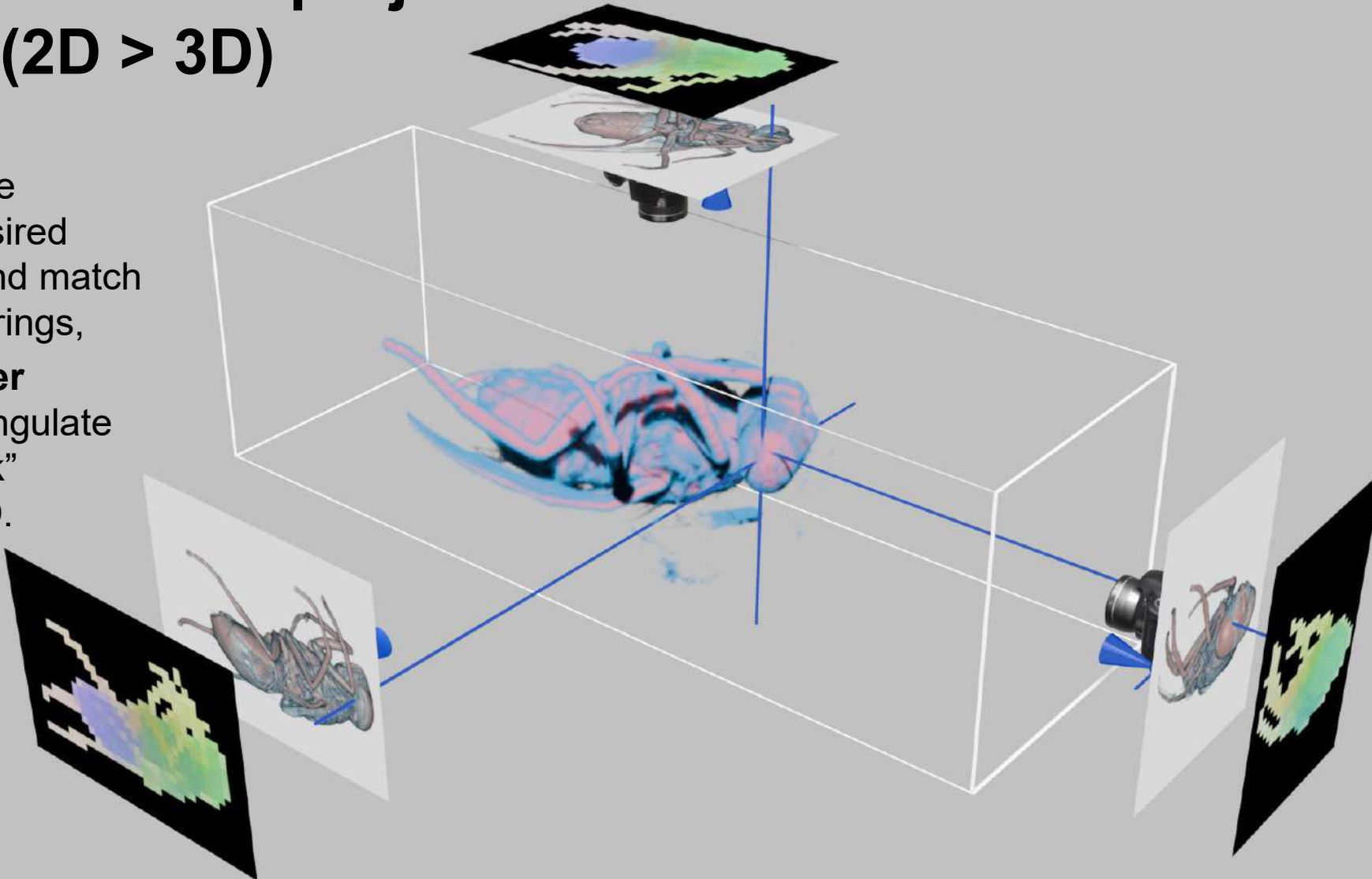
3. Match+project (2D > 3D)

- Pick reference feature of desired “landmark” and match across renderings,
- Use **computer vision** to triangulate the “landmark” position in 3D.



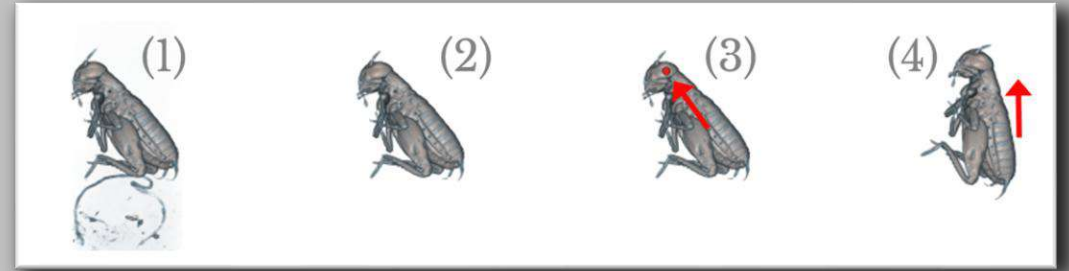
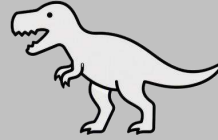
3. Match+project (2D > 3D)

- Pick reference feature of desired “landmark” and match across renderings,
- Use **computer vision** to triangulate the “landmark” position in 3D.





What did we do?

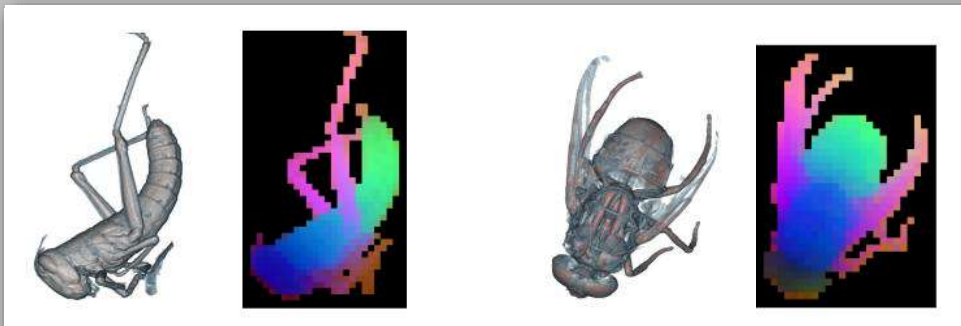


Problem: Estimation of head-body axis and subsequent sample reorientation

Dataset: BugNIST dataset, which contains micro-CT scans of natural heritage samples (bugs) of varied anatomical complexity and initial specimen orientation

Experiments:

- How much image processing is required (segmentation step) to obtain sufficiently good renderings?
- How many viewpoints need to be rendered, to solve the task?
- Must we use the biggest foundation models?
- How precisely can we locate the 3D position of interest points?
- and so on ...



If you want to discuss this or hear more about our experience so far – let's discuss at the poster!