

Towards dense object tracking in a 2D honeybee hive - supplemental material

Katarzyna Bozek

`kasia.bozek@oist.jp`

Laetitia Hebert

`laetitia.hebert@oist.jp`

Alexander S. Mikheyev

`alexander.mikheyev@oist.jp`

Greg J. Stephens

`greg.stephens@oist.jp`

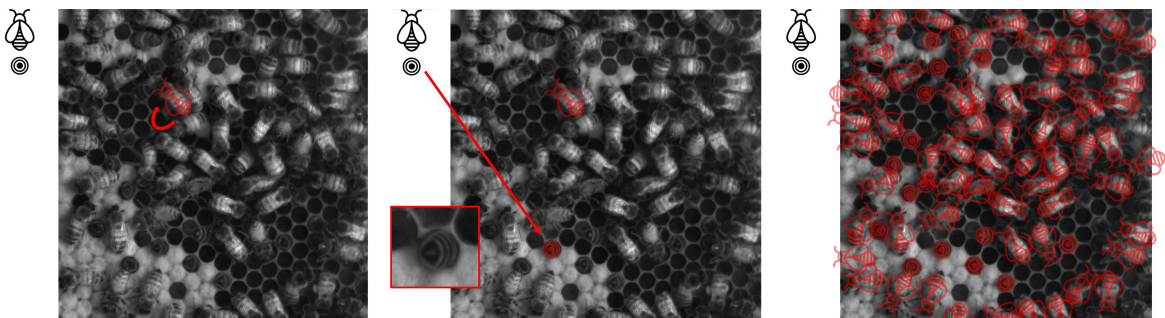


Figure S1: Amazon Mechanical Turk labeling schematic. Annotators were instructed to mark all the bees in an image of a bee comb through dragging and dropping a bee symbol on each bee in the image and matching the symbol's orientation angle. There are 2 symbols for marking - a bee symbol to mark fully visible bees and circle symbol to mark bees that are inside the cells, where only the bee abdomen is visible. Annotators are also instructed to use the same symbols to mark the small number of bees that are upside down.

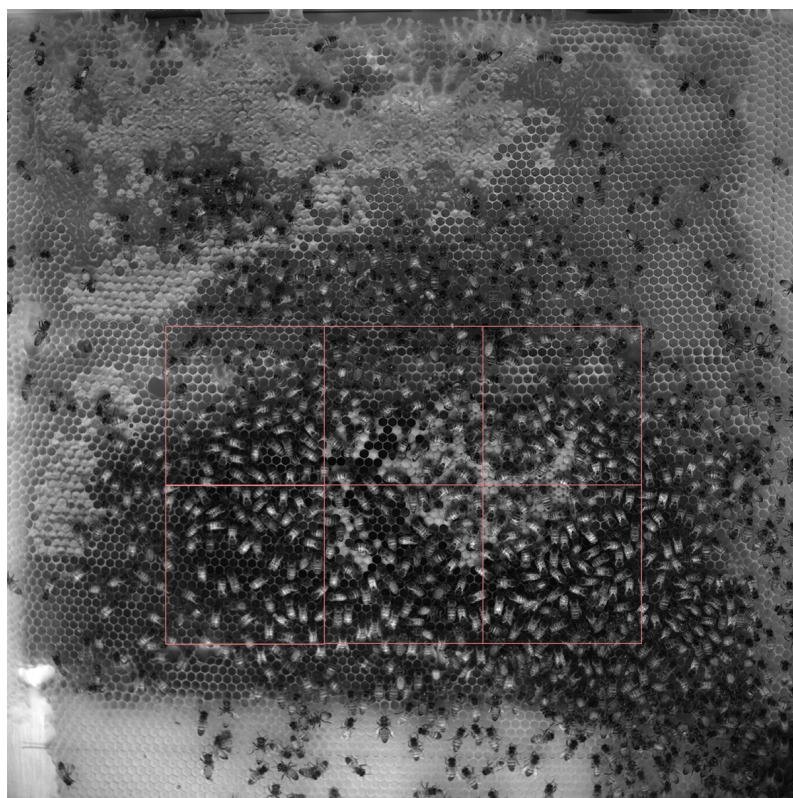


Figure S2: Regions of the 30 FPS beehive video used for human labeling and network training. The squares designate the size of subregions used as one Amazon Mechanical Turk task.

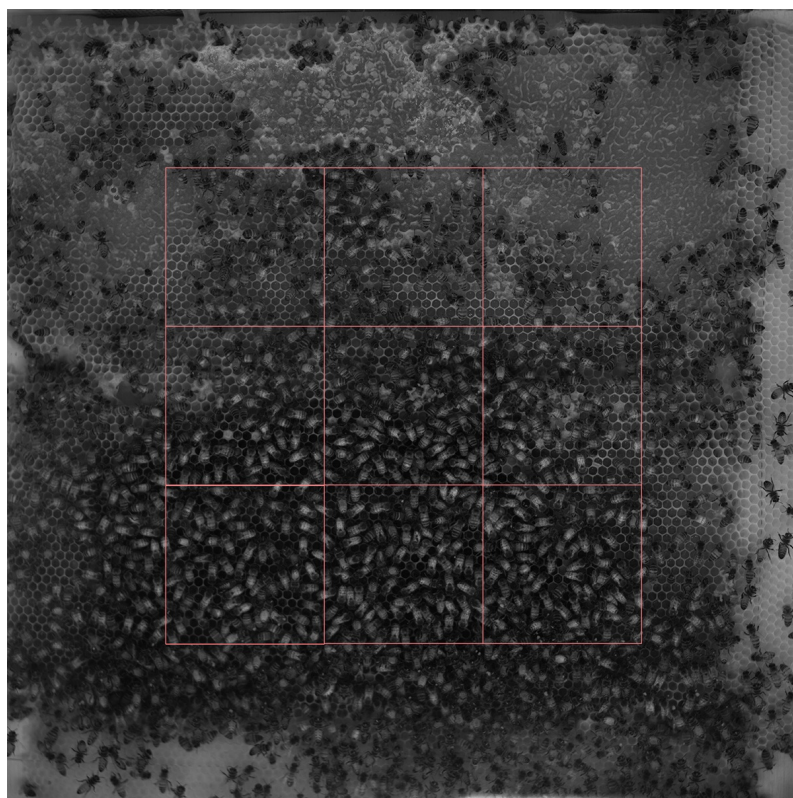


Figure S3: Regions of the 70 FPS beehive video used for human labeling and network training. The squares designate the size of subregions used as one Amazon Mechanical Turk task.

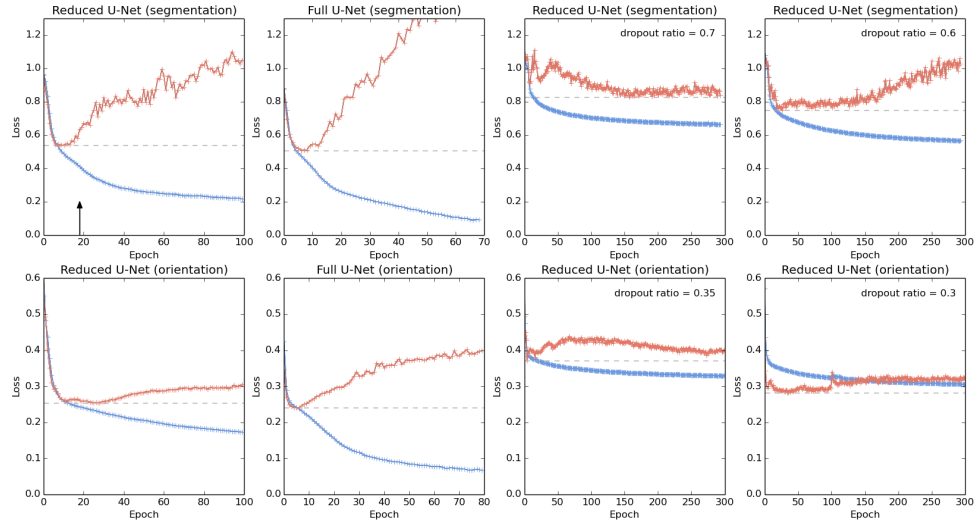


Figure S4: The change of the loss function with training epoch using reduced U-Net (first column), full U-Net (second column) and different levels of dropout (column 3 and 4). The network performing the segmentation task is shown in the first row, the network performing orientation angle search in the second row. The loss for the training set is marked in blue and the loss for the test set in red. The full U-Net results in substantial overfitting, while reducing the size of U-Net reduces the amount of overfitting. Various levels of dropout result in prohibitively slow training (3rd column), or also lead to overfitting with worse overall results (4th column). For this reason we chose early-stop in training (iteration 18 indicated on the upper-left panel) as a measure against overfitting.

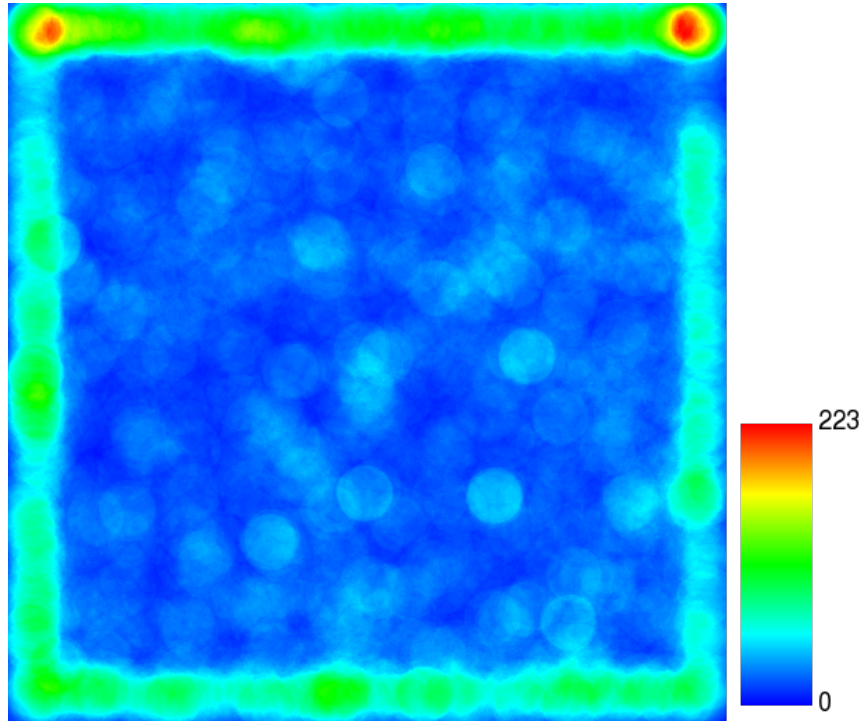
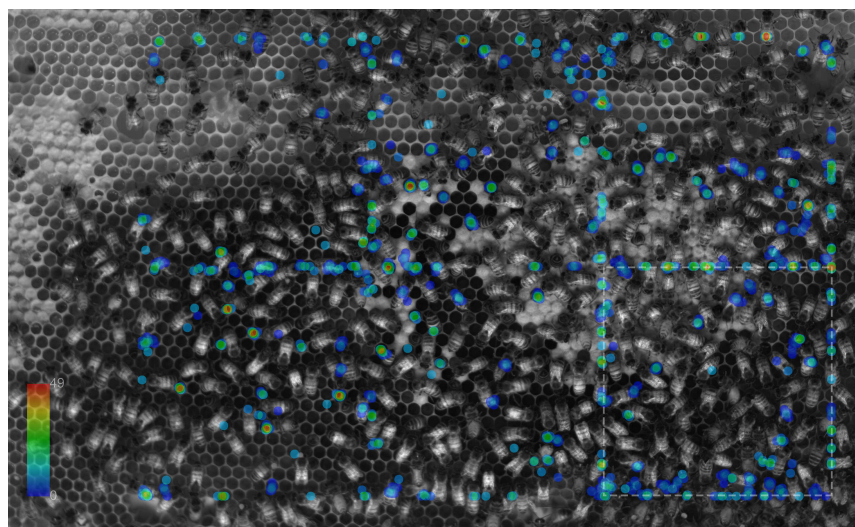
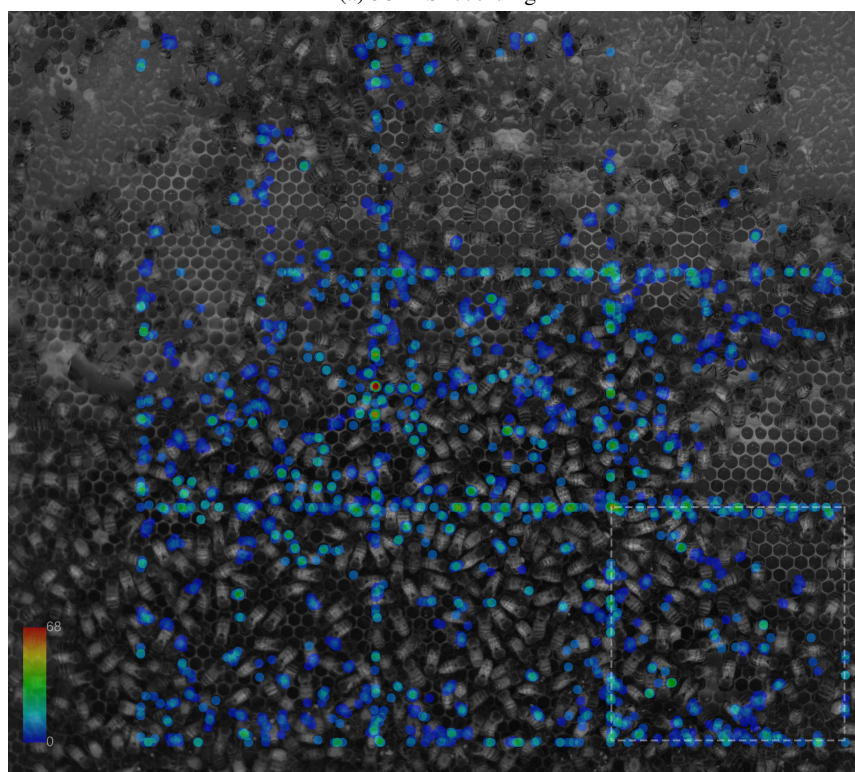


Figure S5: Edge effects of the input image reduce network performance. We show a spatial histogram of the number of incorrectly predicted bees (FP's) across the 512x512 pixel image patches used as input to the network.



(a) 30 FPS recording



(b) 70 FPS recording

Figure S6: Edge effects of the input image increase variability among human annotators. We show a spatial histogram of the location of disagreements in bee labeling among human raters. In an individual AMT task, annotators labeled 1024x1024 pixel image patch, one of which is indicated with a square shape in the images (white dashed outline).

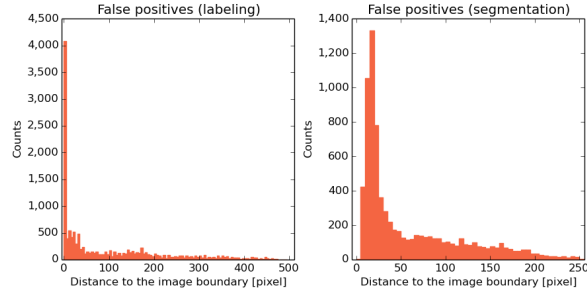


Figure S7: The number of the labeling disagreements among AMT annotators (left) and FP's as identified through the network segmentation (right) is large near the image patch boundary suggesting an edge effect that can be improved in later implementations.

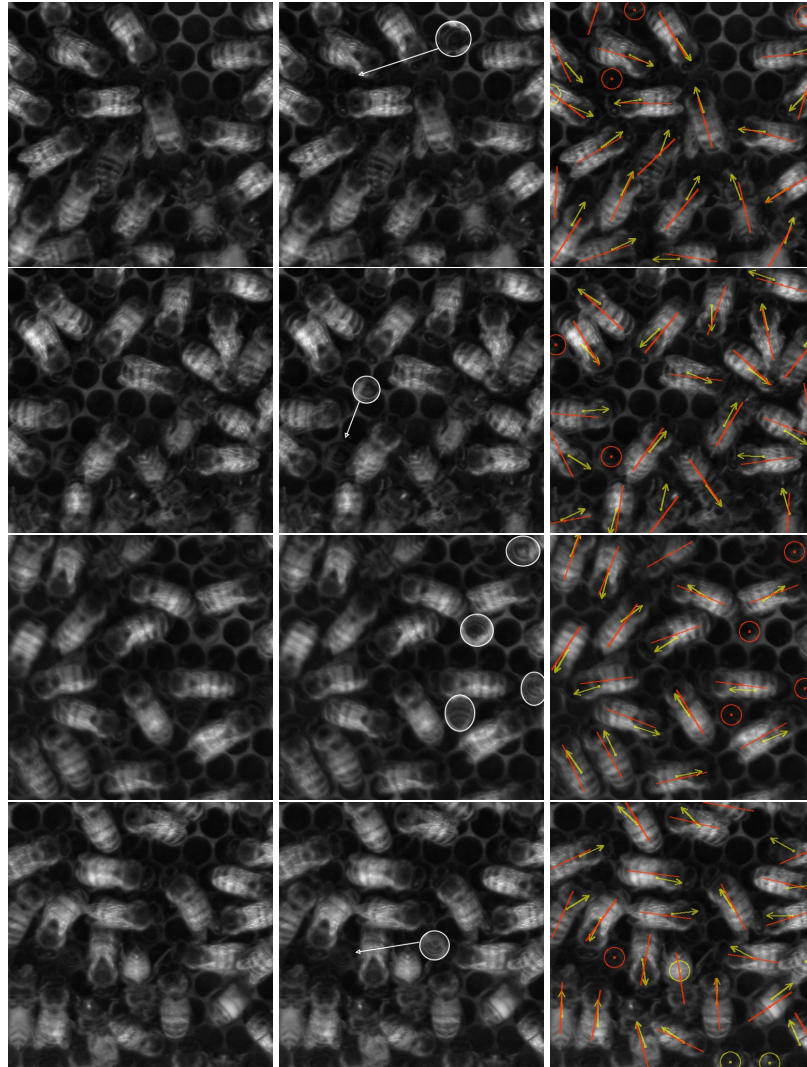


Figure S8: Bees which are partially obscured inside comb cells are often hard to identify by human labelers but still correctly segmented. Shown are examples of difficult to label cases. For the original image (left column), the bee abdomens unnoticed by labelers are highlighted (middle column). These tails were however picked by the segmentation network, which is shown in the corresponding images in the right panel with labels marked in yellow and predictions in red. Such cases contribute to number of FP's in the network performance reported in Table I.

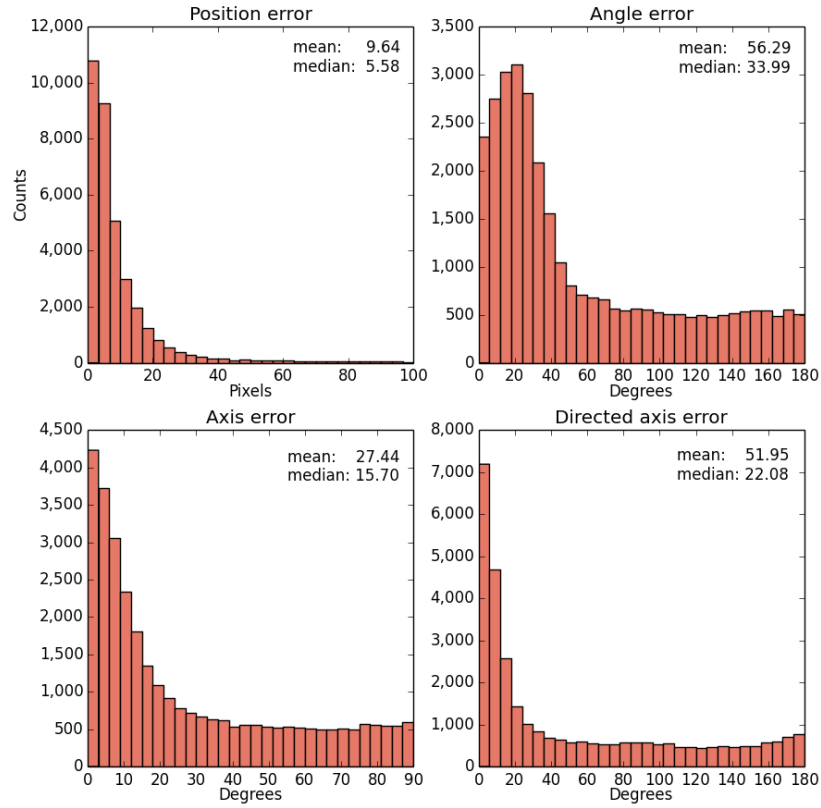


Figure S9: Network prediction errors for all labeled bee instances in the 2,176 images of the test dataset. We show error histograms as well as the mean and median errors for position, orientation angle, axis angle and directed axis angle predictions. The flat tails for angle predictions suggest a small baseline of random predictions.

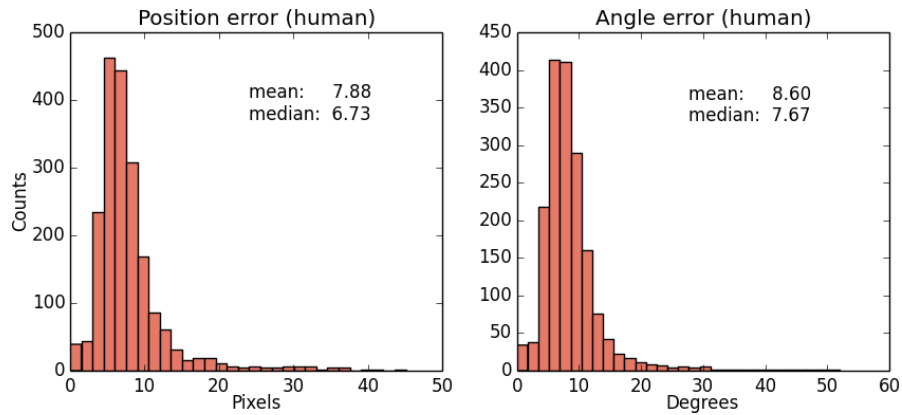


Figure S10: Variability in annotated bee position and orientation among human raters. For 2034 bee instances we show the histogram of the standard deviation of 10 repeated annotation tasks against the one reference annotation used in the dataset for network training and testing.

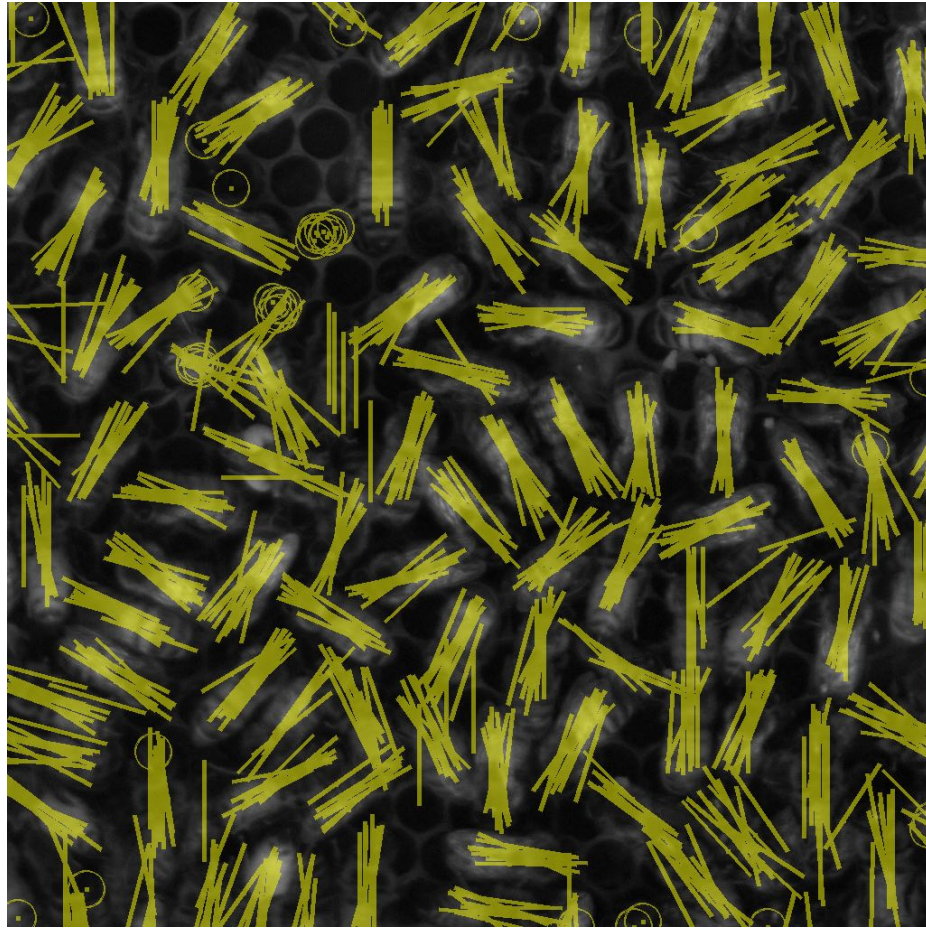


Figure S11: Example of the variability among human raters. Each yellow line is centered and aligned to a honeybee body identified by an annotator from Amazon Mechanical Turk and the same image was presented across 10 different annotators. Circles are centered on image locations identified as bee abdomens.

Supplementary Movie: In the ancillary file “simple_tracking.mp4” we show an example of a reconstructed trajectories. Individuals in one frame are matched to the closest individuals in following frames using position, orientation, angle, and velocity. In case a trajectory is lost, we searched up to five frames ahead for a close match to complete this trajectory.