# BEE POSE ESTIMATION FROM SINGLE IMAGES WITH CONVOLUTIONAL NEURAL NETWORK

Le Duan[1], Minmin Shen[1], Wenjing Gao[2], Song Cui[2] and Oliver Deussen[1]

Interdisciplinary Center for Interative Data Analysis, Modelling and Visual Exploration (INCIDE)[1], University of Konstanz, Germany
Institute of High Performance Computing[2], Singapore

## Abstract

We propose a deep convolutional neural network (ConvNet) based method for detailed bee pose estimation, which aims to detect landmarks as the tips of an bee's antennae and mouthparts from a single image. In this paper, we formulate this problem as inferring a mapping from the appearance of a bee to its corresponding pose. The proposed framework utilizes the powerful representation capability of ConvNet to learn the mapping from the local appearance and global structure of a bee to its corresponding pose. Our method is able to localize a varying number of targets in complex background, especially for the cases when the bee is fed sugar water with a stick. It has been shown that our method outperforms the existing bee pose estimation algorithm on two challenging datasets of bees.

## Introduction

Animal pose estimation is important for behavior study of animals such as bees[1]. Our dataset shows that behaviors of bees could be trained in controlled stimulus conditions, e.g., different light conditions or human interference such as feeding the bee sugar water with a stick(b). These behaviours can be reflected as movements of bee body parts such as their antennae or mouthparts. Bee pose estimation is challenging because the bee body parts exhibit self-similarity and self-occlusions. Moreover, the number of bee body parts could be varying(c-f) and there may be weak correlation among the movements of different bee body parts. To address the aforementioned issues, we present a unified framework that utilizes ConvNets for bee pose estimation.

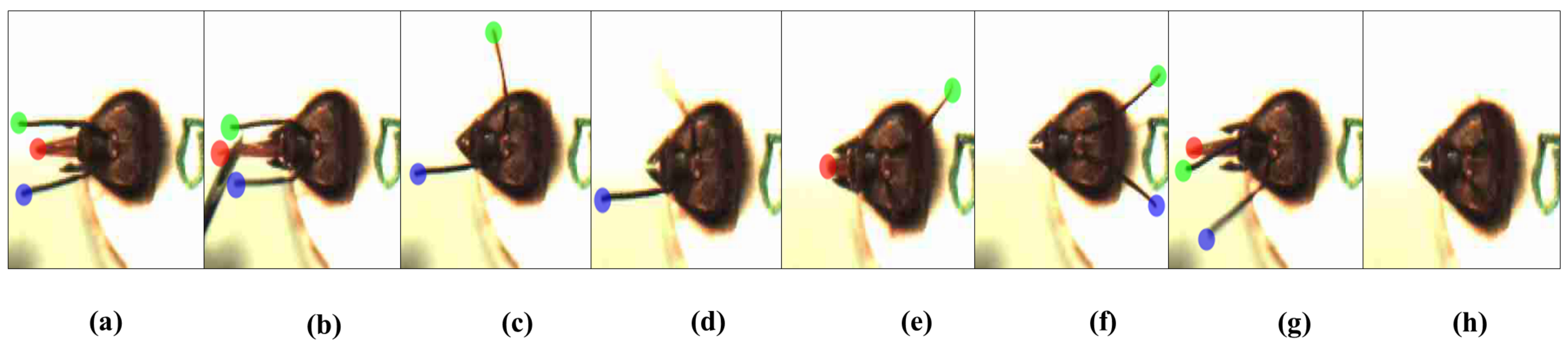


Figure 1 : Example images of various bee poses. The right antenna is represented by green dot, the tongue is represented by red and the left antenna is colored in blue. (a) All tips are present; (b) the sugar water is fed to a bee with a stick; (c)-(e) some body parts are not visible; (f) the antennae may move backwards in some rare cases; (g) part of the tongue is occluded by the right antenna; (h) all parts are absent.

## Pose Estimation Method

We aim to estimate the pose $\mathbf{P} = \{\mathbf{x}^n | 0 \leq n \leq N, \forall \mathbf{x}^n \in \Re^2\}$ from a single image $I$, where $\mathbf{x}^n = \{x^n, y^n\}$ denotes the position of a tip in image coordinate system.

Our framework imposes constraints to the solution space based on two cues: local appearance and global structure to map the image $I$ to the corresponding pose $\mathbf{P}$. On the one hand, we present a new net structure based on VGG-16[2] (Fig. 2b) for predicting confidence map (Fig. 2c) of possible tip positions. On the other hand, we employ a fine-tuned GoogLeNet[3] (Fig. 2d) to extract feature vectors which representing the global structure of input images and construct a feature space. Assuming that similar poses would have similar features, the data point of the test image should lie close to the training images with similar pose in the feature space. We use $K$ nearest neighbour (KNN) search to find the $K$ training images with similar pose as the test image and compute the probability masses of the tips positions (Fig. 2e-f). Combining the local and global information, the pose of a bee can be estimated (Fig. 2g).

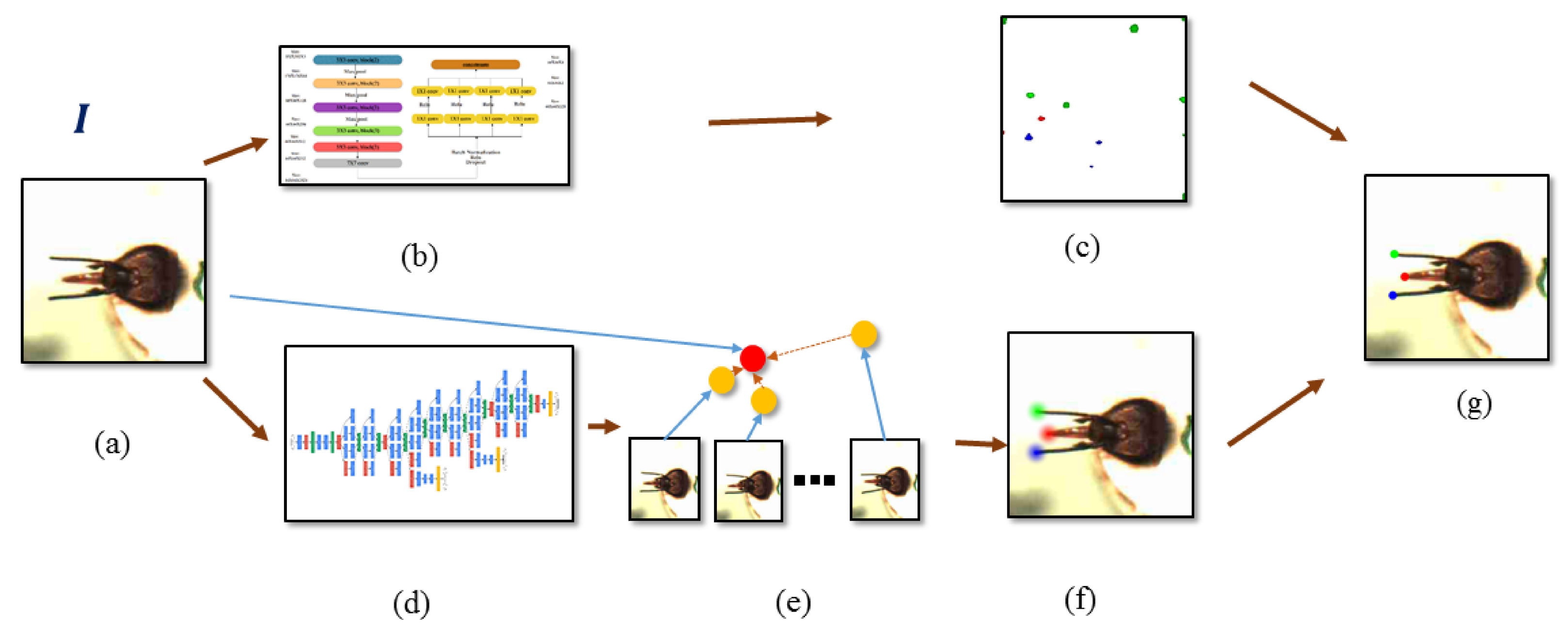


Figure 2 : Data flow.

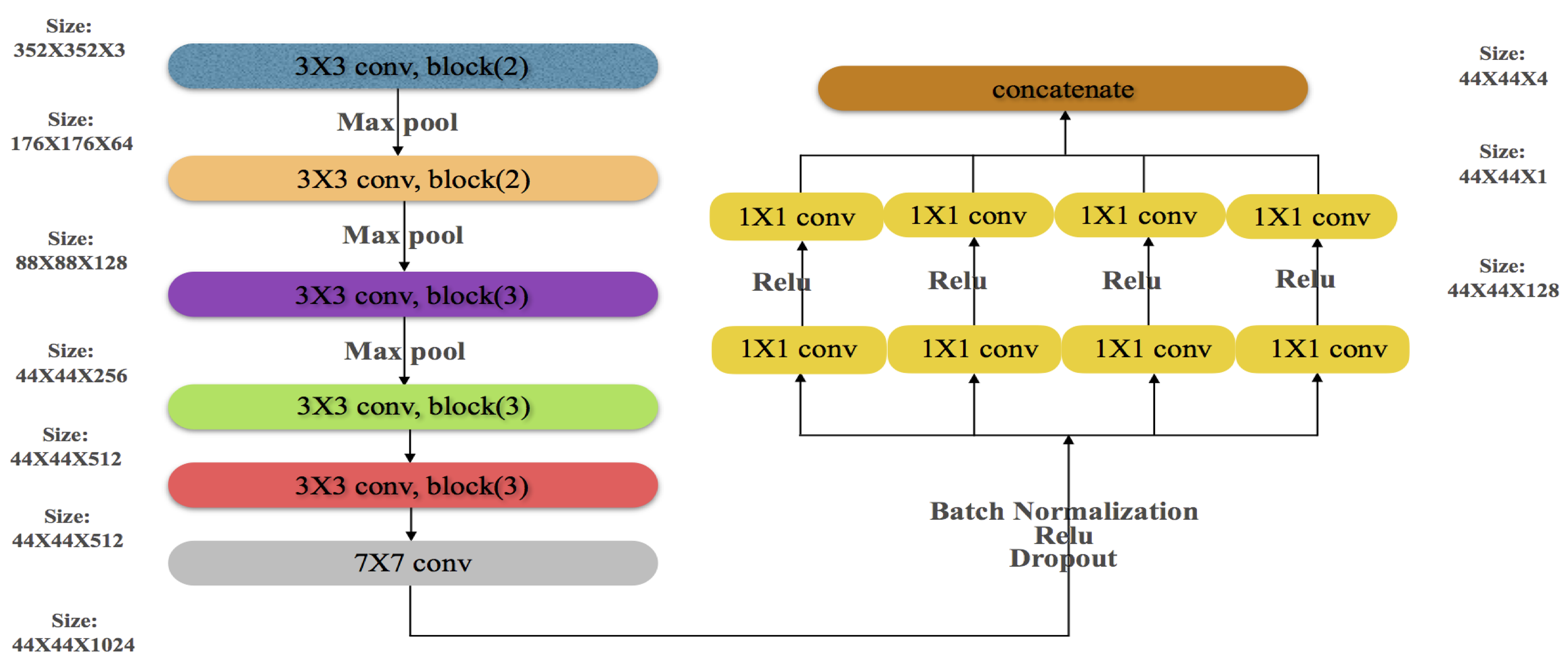


Figure 3 : Details of the modified VGGNet in our framework. All the tips share weights up to 7×7 layer, features for specific tips and background are learned from corresponding path.

## References

[1] Shen M, Szyszka P, Galizia C G, et al. Automatic framework for tracking honeybee's antennae and mouthparts from low framerate video[C]//Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE, 2013: 4112-4116.

[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[3] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[4] Shen M, Duan L, Deussen O. Single-Image Insect Pose Estimation by Graph Based Geometric Models and Random Forests[C]//European Conference on Computer Vision. Springer International Publishing, 2016: 217-230.

## Experimental Results

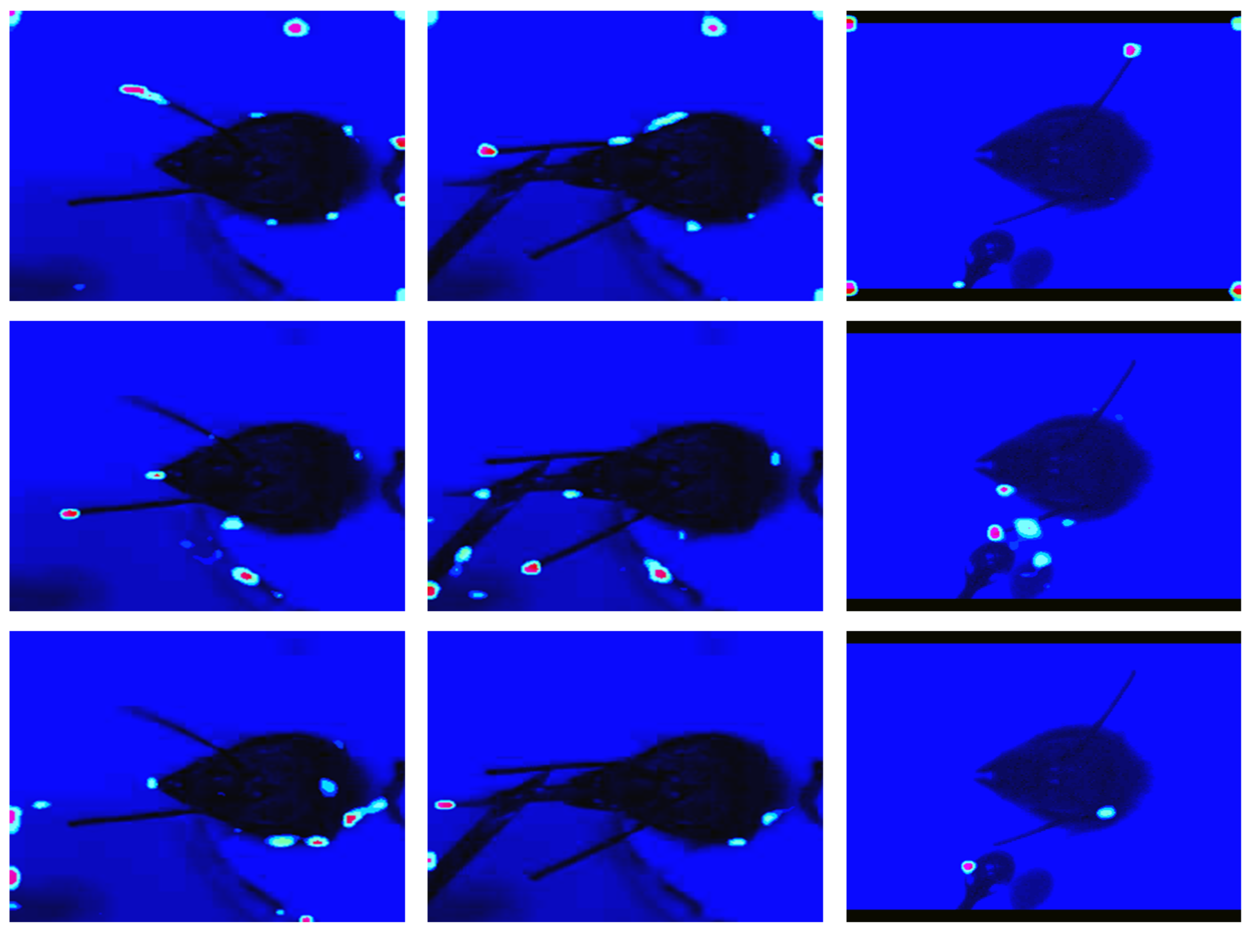

Figure 4 : Examples of confidence maps of different tips.

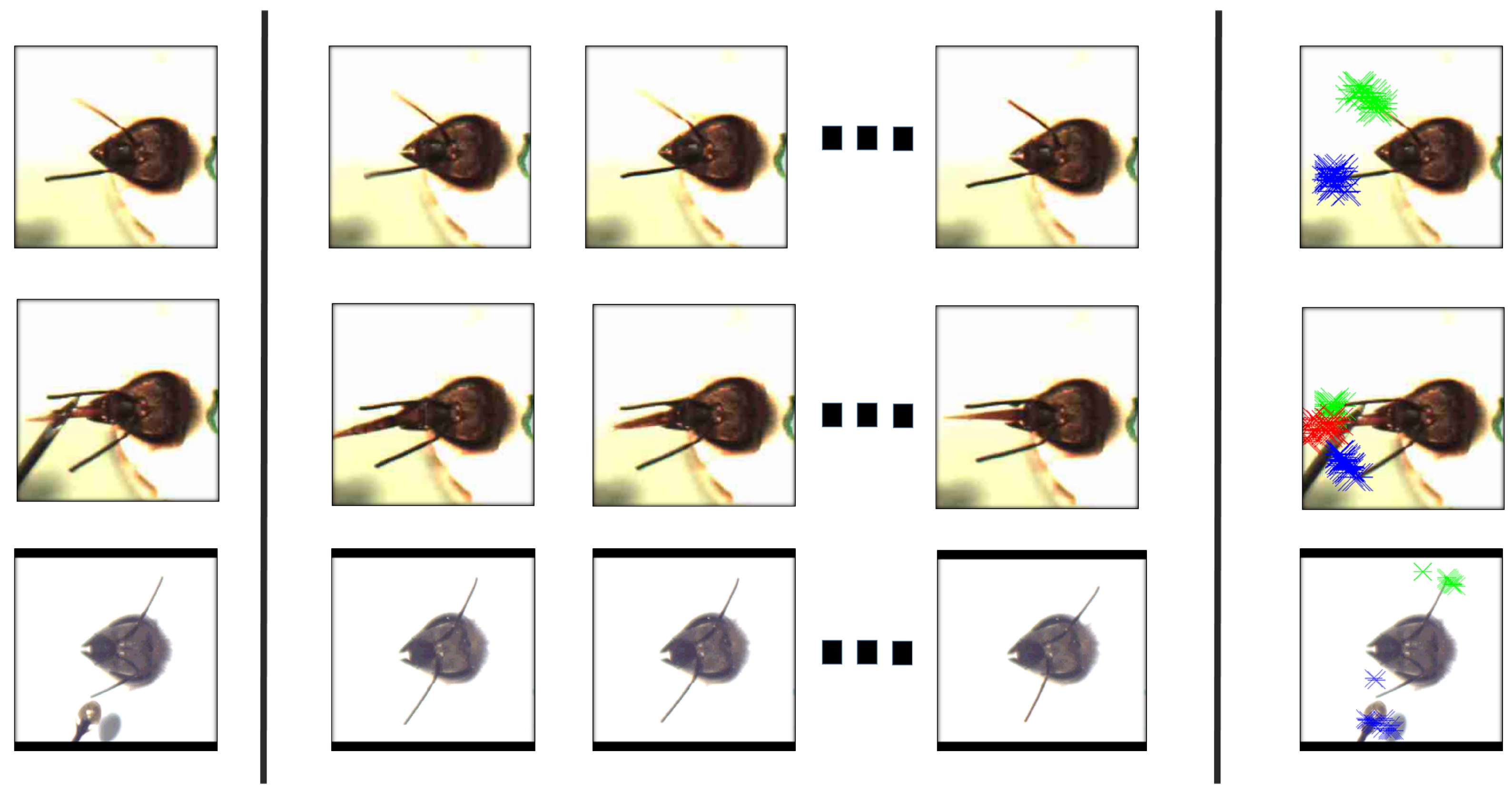

Figure 5 : Examples of KNN results.

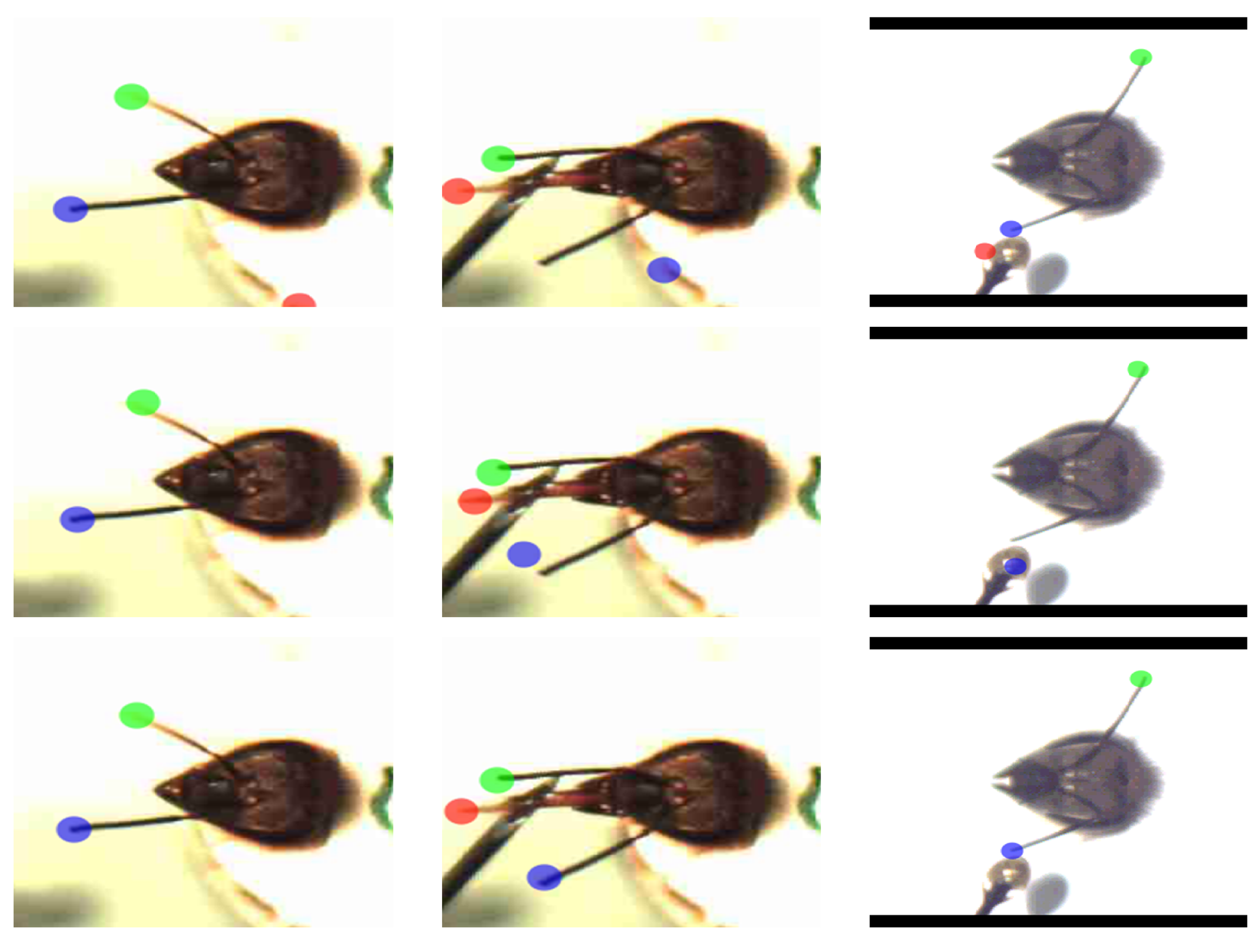

Figure 6 : Qualitative results of using confidence map only (the first row), using KNN only (the second row) and our method (the combination of confidence map and KNN, the last row).

We use three numerical measures, position error in pixel, false positives (FP) and false negatives (FN) to evaluate the experiment results. Position error indicates Euclidean distance between our results and ground truth. We compare the performance of our method and the RF-based method[4] on two datasets. As the images in Dataset B contain no tongue, we only compare the results for two antennae. Table 1 shows the comparison results of two methods on Dataset A and Table 2 shows the numerical comparison of the two methods on Dataset B.

Table 1 : Performance comparison on Dataset A.

| Algorithms | Left antenna | | | Tongue | | | Right antenna | | |
|---|---|---|---|---|---|---|---|---|---|
| | pos. error | FN(%) | FP(%) | pos. error | FN(%) | FP(%) | pos. error | FN(%) | FP(%) |
| Proposed | 5.6 | 3 | 0 | 7.9 | 22 | 2 | 4.3 | 1 | 0 |
| RF-based method | 13.8 | 5 | 0 | 13.6 | 8 | 23 | 8.5 | 5 | 0 |

Table 2 : Performance comparison on Dataset B.

| Algorithms | Left antenna | | | Right antenna | | |
|---|---|---|---|---|---|---|
| | pos. error | FN(%) | FP(%) | pos. error | FN(%) | FP(%) |
| Proposed | 10.2 | 4 | 1 | 7.1 | 4 | 2 |
| RF-based method | 14.3 | 2 | 1 | 7.0 | 4 | 2 |

e-mail:duan.le@uni-konstanz.de, songcui@acm.org