Karol Strama*, Katharina Steeg, Dominik Rzepka, Artur Kos, Gabriele A. Krombach, Katarzyna Heryan, and Michael H. Friebe

Image Segmentation for Automatic Needle Puncture Annotation in Vibroacoustic Signals

https://doi.org/10.1515/cdbme-2025-0176

Abstract: Manual annotation of biomedical data is often a prolonged and error-prone process, particularly in scenarios involving dynamic physical interactions, such as needle insertions into soft tissue. In this study, we present an automated annotation framework using the Segment Anything Model (SAM) to detect puncture events in vibroacoustic recordings from needle insertions into preserved *Manduca sexta* specimens. The annotation tool utilizes SAM to extract segmentation masks from video recordings of needle insertions and analyzes vertical deformations of the resulting mask boundaries. Two puncture moments, one at entry and one at exit, were estimated by searching for minima in the bottom boundary displacements, yielding more reliable results compared to the top boundary. The tool detected the first and second punctures in 85% and 90% out of 300 cases, respectively. Although the tool provides promising accuracy, further refinement could be done to achieve higher and more robust precision. Limitations like background interference, reflections, and needle inclusion can be mitigated by improving the setup or optimizing segmentation models, including fine-tuning SAM.

Keywords: Image Segmentation, Segment Anything Model, Automated Annotation, Vibroacoustic Signals, *Manduca sexta*

1 Motivation

Recent advancements in machine learning have sparked interest in exploring its potential applications across various fields. A crucial part of reliably utilizing machine learning techniques is to prepare an vast and representative dataset and establish a reliable ground truth. In many cases, manual data preparation can be tedious and time-consuming. Several studies developed automatic or semi-automatic annotation methods with varying levels of human involvement across different data types, including video, images, text, and audio recordings [1–3]. For example, Uijlings *et al.* employed a collaborative assistant for semi-automated segmentation mask annotation in images, improving annotation speed by approximately 80% compared to manual annotation and 17% over other state-of-the-art methods [2]. A similar study developed an interactive annotation framework to track objects in videos with segmentation masks inside user-defined bounding boxes, but that required initial user-annotation for every video [3]. Although the aforementioned tools gave promising results, their need to receive external input made them time-consuming, creating the demand for approaches that require less manual intervention.

A comparable need occurred in the investigation of vibroacoustic signals recorded during needle insertions to analyze soft tissue transitions. Vibroacoustic signals originate from needle-tissue interactions and encode for important puncture events, such as entering a cavity or tissue characteristics [4, 5]. Analyzing these data requires a viable annotation to correlate needle movement with signal events. Initially, punctures were annotated manually by marking the video frames close to needle entry or exit, based on visual cues in alongside captured videos. However, when using an *in vivo* model, deformations before needle penetration occurred, causing discrepancies between visual annotations and the actual puncture events (Fig. 1). This led to inaccurate manual annotations, as the visible part of the needle did not correspond to the precise moment of puncture.

To avoid ambiguous annotations, puncture events were detected automatically using the Segment Anything Model (SAM) from Meta, introduced in 2024 [6]. This paper presents an automated tool that leverages SAM as a state-of-the-art image segmentation method to capture dynamic changes during needle insertions.

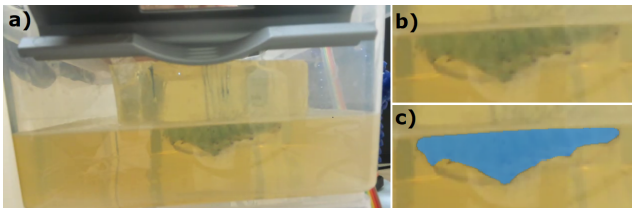

Fig. 1: Sample frame from the recordings (a) as well as the corresponding ROI (b) and segmentation mask (in blue) created by SAM (c).

***Corresponding author: Karol Strama,** Faculty of Computer Science, AGH University of Kraków, Al. Mickiewicza 30, Kraków, 30-059, Poland, e-mail: karol.strama.pl@gmail.com
Katharina Steeg, Gabriele A. Krombach, Faculty of Medicine, Radiology, University Hospital Gießen, Justus-Liebig-University, Klinikstr. 33, Giessen, 35392, Germany
Dominik Rzepka, Artur Kos, Katarzyna Heryan, Michael H. Friebe, Faculty of Computer Science, AGH University of Kraków, Al. Mickiewicza 30, Kraków, 30-059, Poland

2 Experimental setup and methods

An automatic annotation tool was designed to precisely identify the video frame corresponding to a needle puncture into an object.

The particular object punctured was a preserved specimen of *Manduca sexta*, the larval stage of the tobacco hornworm. These insects, approximately 10 cm long and 2 cm in diameter, are well-characterized model organisms in biomedical research, providing an emerging platform for *in vivo* studies. A freshly preserved larva was placed inside a gelatin block, with a small section removed beforehand to accommodate the specimen. A second gelatin block was placed on top to stabilize the needle during insertion. Gelatin was chosen for its transparent properties, allowing for parallel video recording of the punctures as the needle advanced for later annotations.

All needle insertion experiments were captured by a video camera mounted at a fixed angle parallel to the gelatin layer (Fig. 1).

The specimens were found to exhibit vast deformations at the needle entry and exit due to their elasticity, which motivated the use of computer vision techniques to establish the automated annotation tool. A versatile Segment Anything Model (SAM) [6] was used to automatically extract the shape of the specimen for each frame of the videos. The model was chosen for the fact that no additional fine-tuning is required to perform well on unseen objects. Furthermore, Kirillov *et al.* report that the model exhibits excellent zero-shot performance that achieves results comparable to methods based on supervised learning [6]. Due to the standardized video recordings, the larva could always be located in a user-defined cropped frame, accelerating the model's computational inference. The corresponding region of interest was thoroughly selected to include the specimen in all available recordings. Subsequently, two SAM methods of principal mask extraction were investigated: (1) generating the mask using a fixed annotation point and (2) generating the mask within a fixed annotation bounding box. It was found that the bounding box method gave more robust results, hence it was selected for all further analysis. An exemplary maximum score segmentation mask along with the image it was generated from is presented in Fig. 1.

To quantify the changes in the generated masks, only their contours were analyzed. Given the horizontal orientation of the specimens and the significant deformations occurring along the vertical needle insertion path, the top and bottom sections of the contours were selected for simplicity.

3 Processing of segmentation masks

The vertical positions of the specimen's top and bottom boundaries were analyzed over time in order to find the exact moments of the needle entry and exit (Fig. 2). Since the larva experiences peak tensile deformation just before the rupture and rapid tissue relaxation thereafter, it was expected that the lowest vertical boundary locations would strongly correlate with the puncture times. Thus, the purpose of the analysis was to search for the minima over time of the obtained segmentation contour coordinates. Investigation of the time evolutions of the mask boundaries showed that values beyond the 150th pixel exhibited erroneous behavior (Fig 2). Inspecting the segmentation masks, we verified that the SAM predictions had consistent errors in that area for all recordings. Hence, a subregion was selected that discarded the pixels from the 150th onward and below the 20th due to the inaccurate edge estimations.

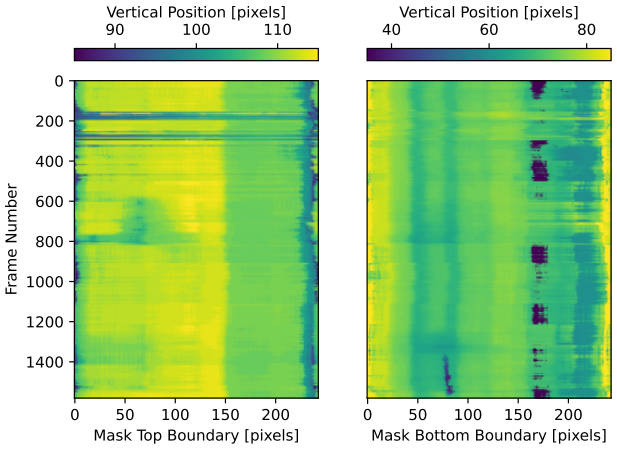

Fig. 2: Time evolution of the top and bottom boundaries as functions of the frame number. Erroneous pixels beyond the 150th were excluded, as no major deformations occurred in the isolated area, but rather the larva moves holistically.

The resulting boundary deformation graphs were used to determine puncture moments. Initially, the minimum boundary value was considered for each frame, as it theoretically tracked the needle tip's contact point. However, this approach proved unreliable due to erroneous SAM predictions, including instances where the model incorporated the needle into the mask, leading to the minimum value not increasing after the puncture.

To address these issues, the mean boundary value was chosen instead, as it significantly decreased the errors. Following this, a mean filter of length 10 pixels was applied to smoothen the curve and exclude outliers, ensuring more robust minima detection (Fig. 3).

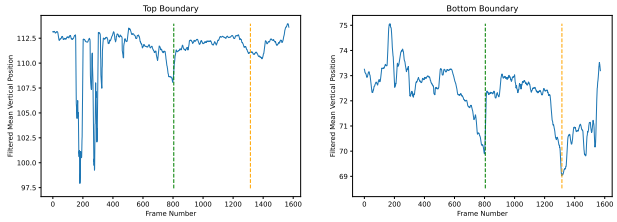

Fig. 3: Filtered mean vertical position movement over time for the top and bottom boundaries. The first and second puncture moments at frames 805 and 1315 have been highlighted with green and orange lines, respectively. The bottom boundary slowly decreased until the point of puncture, and then quickly increased before stabilizing after each puncture. The slower rise after the second puncture occurred due to the needle tip contained within the mask which shifted the average down.

The top boundary results contained significant errors, even after filtering. A longer mean filter could have been applied, but it would also shift other parts of the functions, leading to less accurate minimum. Similar behavior of the top boundary deformations was found for other recordings. In contrast, the bottom boundary gave more promising results with fewer dominant outliers.

The bottom boundary analysis yielded much more robust estimations for both puncture times. Although the top boundary produced a local minimum around the first puncture, it was difficult to extract since it was not a global minimum (Fig. 3). Therefore, a consistent structure of the automated annotation algorithm was selected. First, the global minimum of the filtered mean vertical position of the bottom boundary is determined and treated as the second puncture moment. The algorithm then searches for the most prominent local minimum prior to this moment and selects it as the first puncture time.

4 Results

The performance of the tool was tested on a dataset of 300 recordings, and the automated estimations were compared with those obtained manually. The manual annotation process required pausing the recordings during punctures and noting the corresponding frame numbers. The automatically gathered annotations experienced a systematic shift in time of roughly 20 to 30 frames relative to the reference manual annotations, most certainly due to human reaction time. With the discarded reaction time bias, the second puncture results for manual and automated annotations differed by at most 20 frames 90% of the time. In the remaining 10%, the estimates varied more significantly, so these were treated as incorrect. Hence, automatic annotations were considered valid if they deviated from the manual reference annotations by no more than 40 frames. On the other hand, the first puncture moment was correctly determined 85% of the time. Table 1 presents the algorithm accuracies if different maximum shifts were selected for the criterion. For most of the recordings, the annotations were robustly determined because of significant global minima. However, for some measurements, the local minima were not as prominent and therefore prone to errors. Lastly, it was observed that for 10% of the time, an erroneous noticeable local minimum was present between the two correct ones, which would have led to incorrect annotations if it had been slightly more significant.

	Maximum Shift			
	40	**30**	**20**	**10**
First Puncture	**85%**	**80%**	**60%**	**35%**
Second Puncture	**90%**	**75%**	**60%**	**45%**

Tab. 1: Percentage accuracy of the algorithm for different maximum allowed shifts between the manually and automatically estimated moments of puncture.

5 Discussion

An automated annotation tool using SAM was developed to detect needle punctures by segmenting the specimen and analyzing vertical deformations of the segmentation masks. Puncture moments were identified through analysis of segmentation mask changes in time, achieving 85 and 90% accuracy for the first and second puncture, respectively. The second puncture moment estimation was found to be more precise and robust than the first one. This result was expected since the algorithm based its estimations on the bottom boundary displacements. Furthermore, for many test cases, the tool performed accurately even though it was difficult to determine the first puncture moment manually.

The model inaccuracies were suspected to be mainly caused by the image background as it consisted of many objects with similar colors as *M. sexta*. Objects with comparable color and texture are often difficult to distinguish for the segmentation model, especially on such small bounding boxes with decreased resolution, which is a common issue [7]. Moreover, the room lighting gave rise to several reflections inside the container which further decreased the model accuracy. Another study highlighted the problem of light reflections in transparent materials for image processing and computer vision tasks [8]. This issue was further addressed in the context of image segmentation and a novel reflection removal technique that uses a single image for reflection suppression was suggested [9]. Additionally, the problem with puncture moment estimation using the top boundary might have been caused by having a separation between gelatin layers in the

vicinity of the *M. sexta* top boundary, and hence the difficulty in discerning between the two. This is supported by the complex background discussion [7].

The first and foremost solution to these problems would be to improve the experimental setup. By providing a container without similar background, well-designed room lightning, and more thoughtfully prepared gelatin layers, it would have been easier for SAM to perform reliable segmentation. Furthermore, utilizing cameras with higher resolution as well as choosing optimized contrast filter in video post-processing, would further enhance the tool's precision. Flores *et al.* improved segmentation in a biomedical setting by studying adaptive sigmoidal contrast enhancement [10]. Moreover, a more appropriate metric than the mean and a more advanced filter could also be used to quantify the segmentation results. Lastly, exploration of different segmentation models could be performed. A recent study presented a model based on SAM that enables more accurate image segmentation [11]. As noted before, SAM was not robust in providing fine-grained details which was especially seen beyond the 150th pixel in the estimated masks and during needle inclusion in the masks after puncture. The authors offer a modified approach called dichotomous image segmentation which was specifically designed to deal with high-accuracy segmentation tasks. Ma *et al.* show another segmentation model based on SAM that was implemented on a medical dataset [12]. An equivalent technique could be used for the *M. sexta* segmentation use case where versatile SAM as the base model could be fine-tuned to match with the specific data containing larva images.

6 Conclusion

Image segmentation is a powerful tool that may be used for various tasks such as determining the instance of a well-defined event. The present study showed that under certain conditions SAM may give promising results in the case of *M. sexta* puncture annotation, achieving the accuracy of 85% and 90% for the first and second puncture annotation, respectively. However, such precision is not satisfactory for the annotation tool to be fully functional. Analysis of mask boundaries showed that significant minima or abrupt changes in the vertical location graphs do not always occur. Some potential sources of errors and possible approaches to solve them have been discussed. The most trivial solution would be to improve the experimental setup by ensuring that the background does not merge with the larva and that there are no striking image reflections inside the container since the larva-background interference significantly influenced the generated masks. Additionally, *M. sexta* could be entirely enclosed within gelatin for complete visibility of the top boundary. More advanced techniques could include changing the model from SAM to a one particularly designed for fine-grained mask boundary estimation, or fine-tuning SAM to a specific dataset containing various larva images.

Author Statement

Research funding: This research is a co-funded project of the National Science Centre (NCN), Poland, No. 2021/43/I/ST7/03098, and the Deutsche Forschungsgemeinschaft (DFG), Germany, No. 504923173, under the OPUS-22 (LAP) framework. Ethical approval: This study does not involve human participants or animals that are subject to ethical approval.

References

[1] Wang S, Li C, Wang R, Liu Z, Wang M, Tan H, et al. Annotation-efficient deep learning for automatic medical image segmentation. Nat Commun 2021;12(1):5915.

[2] Uijlings JRR, Andriluka M, Ferrari V. Panoptic image annotation with a collaborative assistant. In: Proc 28th ACM Int Conf Multimedia. 2020. p. 3302–3310.

[3] Chen B, Ling H, Zeng X, Gao J, Xu Z, Fidler S. Scribblebox: Interactive annotation framework for video object segmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16. Springer; 2020. p. 293–310.

[4] Spiller M, Esmaeili N, Sühn T, Boese A, Turial S, Gumbs AA, et al. Enhancing Veress Needle Entry with Proximal Vibroacoustic Sensing for Automatic Identification of Peritoneum Puncture. Diagnostics 2024;14(15):1698.

[5] Serwatka W, Heryan K, Sorysz J, Illanes A, Boese A, Krombach GA, et al. Audio-based tissue classification—preliminary investigation for a needle procedure. Curr Dir Biomed Eng 2023;9(1):347–350.

[6] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. In: Proc IEEE/CVF Int Conf Comput Vis. 2023. p. 4015–4026.

[7] Wang M, Zhang Y. Image Segmentation in Complex Backgrounds using an Improved Generative Adversarial Network. Int J Adv Comput Sci Appl 2024;15(5).

[8] Wan R, Shi B, Duan L-Y, Tan A-H, Gao W, Kot AC. Region-aware reflection removal with unified content and gradient priors. IEEE Trans Image Process 2018;27(6):2927–2941.

[9] Elnenaey A, Torki M. Utilizing Multi-step Loss for Single Image Reflection Removal. arXiv preprint arXiv:2412.08582; 2024.

[10] Flores WG, de Albuquerque Pereira WC. A contrast enhancement method for improving the segmentation of breast lesions on ultrasonography. Comput Biol Med 2017;80:14–23.

[11] Liu X, Fu K, Zhao Q. Promoting Segment Anything Model towards Highly Accurate Dichotomous Image Segmentation. arXiv preprint arXiv:2401.00248; 2023.

[12] Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nat Commun 2024;15(1):654.