Proceedings on Automation in Medical Engineering
Proc AUTOMED, 2026, Vol. 3, No 1, Article ID 2514, 2 Pages

Infinite Science Publishing

Why we need to consider perspective in image-based surgical tool classification

A. Millán Cerezo [1,2*], J. Badilla-Solórzano[3], T. Seel[2], T. S. Rau[1], L. Budde[2]

[1] Hannover Medical School, Department of Otolaryngology and Cluster of Excellence EXC 2177/1 "Hearing4all", Hannover, Germany
[2] Leibniz University of Hannover, Institute of Mechatronic Systems, Hannover, Germany
[3] National University of Costa Rica, Department of electrical engineering, San José, Costa Rica.
[*] Corresponding author, email: millancerezo.anais@mh-hannover.de

Abstract: Automated surgical assistance systems rely on reliable instrument detection, yet current deep learning models remain sensitive to visual ambiguity. One overlooked factor contributing to this problem is perspective, which can hide or reveal critical discriminative features. We introduce a perspective-aware evaluation framework using synthetic data to analyze how perspective influences classification performance. Our results show that detection accuracy varies with viewpoint, identifying both optimal and ambiguous perspectives. These findings suggest that incorporating perspective-awareness may improve robustness in automated surgical systems.

I. Introduction

The current shortage of trained nursing personnel has accelerated research into automated systems for surgical assistance [1][2]. In this context, a robotic scrub nurse (RSN) is expected to autonomously detect and handle surgical instruments. Although recent advances in deep learning (DL) have significantly improved marker-less, image-based instrument detection under controlled conditions, performance still falls short of human perception. In practical scenarios, instruments may differ only by subtle visual features, be partially occluded or overlap with other tools, resulting in ambiguous visual information [3]. A key contributing to this gap lies in how humans and DL models interpret visual information.

Humans are highly skilled at interpreting visual scenes, even when information is incomplete or ambiguous. We instinctively recognize when visual information is missing and seek additional context. A key example of this behavior is perspective: certain viewpoints may withhold essential features, while others reveal them. When ambiguity arises, humans naturally change their perspective or context to resolve uncertainty.

This raises an important question: can perspective-dependent performance be leveraged to improve detection accuracy in RSN systems? More specifically: can models be designed to recognize uninformative viewpoints and the need for additional visual context, similar to human visual reasoning?

In this work, we investigate how perspective influences image-based classification of surgical instruments. Using retractors as a case study (Figure 1), we provide a systematic analysis of viewpoint-dependent model performance. This aims to identify when a given view is insufficient to generate reliable predictions and to inform how additional visual information could improve robustness in detection systems. Our findings suggest that incorporating perspective-awareness into detection systems may be a key step toward more reliable and context-aware detection systems in automated surgical assistance.

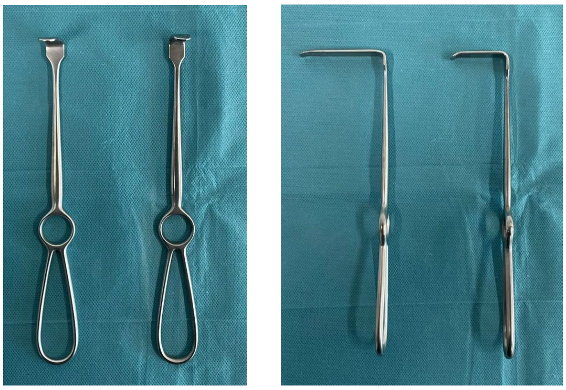

Fig. 1 Long and short retractors seen from two perspectives

II. Materials and Methods

Two slightly different surgical retractors were chosen as a subject for the perspective analysis. These are particularly suitable, as the blade can easily be revealed or obscured depending on the viewing angle. A pipeline was developed for the

II.I Perspective-aware data generation

We generated perspective-aware synthetic data using 3D models of the instruments in a virtual simulation environment (Blender v 4.4) to investigate the effect of perspective on the model performance. A local coordinate system was defined for each instrument, and images were rendered using a virtual camera positioned in a spherical coordinate system (see Figure 2). This setup enabled systematic sampling across a viewing hemisphere while maintaining full control over the orientation of each instrument relative to the camera.

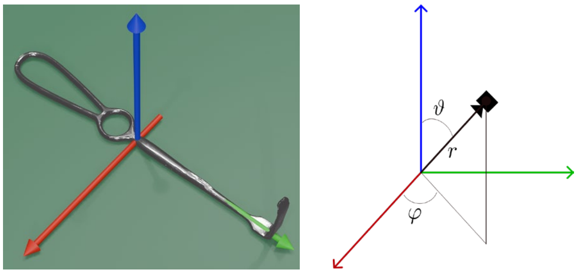

Fig .2 (a) 3D model of large retractor with defined coordinate system (b) Spherical CS defined by r, φ and θ

II.II Evaluation methodology

A total of 20.000 images were generated and split 80/20 into training and validation sets to train a YOLO11-large model [4]. For the test dataset, 4.500 samples associated with spherical coordinates were used, enabling direct extraction of perspective information during evaluation. Model predictions were analyzed depending on the radius (*r*), azimuthal (*φ*) and polar (*θ*) angles by computing performance across the viewing hemisphere. Finally, perspective heatmaps for the accuracy and precision were generated for each class, illustrating viewpoints that enable reliable recognition as well as perspectives that introduce ambiguity.

III. Results and Discussion

The proposed pipeline enabled a controlled synthetic data generation process, allowing the creation of balanced datasets and precise pose tracking for each instrument relative to the camera. This allowed the evaluation of detection performance explicitly as a function of perspective.

Rather than summarizing performance with a single scalar metric, the resulting perspective heatmaps, provide a spatial representation of model behavior across the viewing sphere (Figure 3). These visualizations highlight regions where detection consistently succeeds, as well as perspectives associated with ambiguity or misclassification.

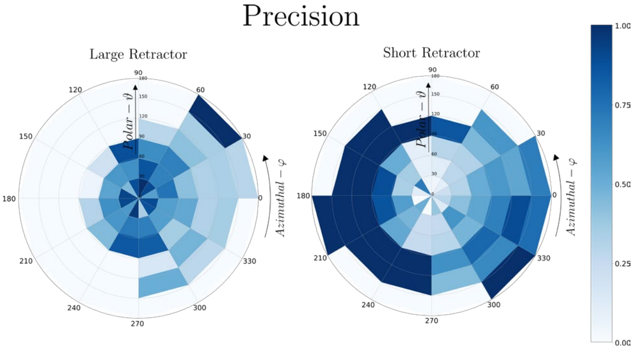

Fig. 3 Perspective-based precision for the retractors

Figure 3 presents precision heatmaps for the retractor instruments, revealing a strong dependence of detection performance on perspective. The short retractor is more reliably detected from side views, where the length and shape of the blade are clearly visible, while other viewpoints lead to increased ambiguity.

These results were obtained in a controlled setting focusing exclusively on retractors, allowing for a clear analysis of perspective-dependent behavior within a single instrument category. How the presence of additional instrument types influences these patterns remains an open question, as inter-class similarities may further amplify viewpoint-dependent ambiguity.

Beyond analysis, the extracted perspective-dependent performance information could inform adaptive camera strategies. For instance, a system could adjust its viewpoint when operating in ambiguous regions, forming the basis for a perspective-aware detection system. Such a strategy can complement stereo approaches, enabling systems to actively select perspectives that resolve ambiguity. However, since the pipeline relies on the relative orientation between the instrument and the camera, practical deployment would require either knowledge of likely orientation distributions or the ability to estimate instrument pose at inference time.

V. Conclusion

Perspective is often overlooked as a factor influencing the performance of deep learning–based object detection. In this work, we introduced a pipeline to systematically study how viewing perspective affects detection performance. Using two visually similar retractors as a case study, we demonstrated that perspective plays a critical role in distinguishing between instruments. At the same time, the analysis was limited to a small set of instruments, and how the presence of additional instrument types, occlusions, or overlaps influences these performance patterns remains an open question.

Beyond this initial analysis, the results motivate future extensions toward perspective-aware decision-making. Multi-view strategies based on optimal or complementary viewpoints could be used to reduce misclassification by actively resolving ambiguity. More broadly, perspective-aware evaluation may support pose-aware classification models and intelligent viewpoint selection in robotic or automated imaging systems. Together, these directions highlight the potential of perspective-awareness as a step toward more robust and context-aware visual perception systems.

Author's Statement

Research funding: The authors state no funding involved. Conflict of interest: The authors state no conflict of interest

References

[1] Harms, Peter. Nursing: A critical profession in a perilous time. Industrial and Organizational Psychology, 2021, 10.1017/iop.2021.58.

[2] Kyrarini M, I. et al. F. A Survey of Robots in Healthcare. *Technologies*, 2021, https://doi.org/10.3390/technologies9010008

[3] Badilla-Solórzano, Jorge, I. et al. (2023). Improving instrument detection for a robotic scrub nurse using multi-view voting. International journal of computer assisted radiology and surgery. 2023, 10.1007/s11548-023-03002-0.

[4] Jocher, Glenn, I. et al.: Ultralytics YOLO11. https://github.com/ ultralytics/ultralytics. Version: 2024