# Detection of Surgical Instruments Using YOLO: An Enhanced Approach

Shubham Laxmikant Deshmukh
shubhamd23@vt.edu
Virginia Polytechnic Institute and State University
Falls Church, Virginia, USA

Pradyumna Kombethota Ramgopal
pradyumna@vt.edu
Virginia Polytechnic Institute and State University
Falls Church, Virginia, USA

## Abstract

The growing application of computer vision in the field of medicine has led to significant improvements in surgical assistance and automated monitoring. The accurate detection of the instruments during surgical procedures is a crucial task for enhancing operational efficiency and patient safety. Previously object detection methods like RetinaNet, have been utilized for surgical instrument detection, but our approach using YOLO is expected to achieve a higher mean Average Precision (mAP). The current work presents an enhanced detection pipeline, where YOLO is fine-tuned on a neurosurgical instrument dataset, achieving superior mAP when compared to prior models. .

***Keywords:*** Deep Learning, Object Detection, YOLO, mAP.

**ACM Reference Format:**
Shubham Laxmikant Deshmukh and Pradyumna Kombethota Ramgopal. 2018. Detection of Surgical Instruments Using YOLO: An Enhanced Approach. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 Introduction and Related Work

Deep Learning has seen significant advancements over the past decade, especially with the advent of architectures such as Convolutional Neural Networks (CNNs) for image recognition and processing tasks. These advancements have enabled breakthroughs in computer vision, leading to widespread adoption in various domains, including healthcare. One particularly important application of deep learning in healthcare is the detection and classification of surgical instruments during operations. Accurate and efficient detection is critical for ensuring patient safety, enhancing operational efficiency, and minimizing the risk of complications during surgeries.

The motivation for detecting surgical instruments stems from the need to automate parts of surgical workflows, enabling real-time feedback to surgeons, improving tool tracking, and assisting in robotic surgeries. While conventional methods, including feature-based and machine learning approaches, have been employed, they often fall short in terms of accuracy and speed. This is particularly the case when detecting multiple instruments simultaneously or when dealing with complex surgical environments. Modern object detection frameworks, like YOLO (You Only Look Once), have shown great promise in overcoming these limitations, offering real-time processing and high detection accuracy.

Our work is primarily focused on detecting neurosurgical instruments in real-time using the Simulated Outcomes Following Carotid Artery Laceration (SOCAL) dataset, which provides a wide range of annotated images depicting surgical tools used during neurosurgery [1]. The dataset contains 365 trials from 177 surgeons, with 31,443 annotated frames from 147 video trials. It includes detailed annotations for eight surgical tools, such as suction, grasper, cottonoid, and muscle, across varied conditions. Each frame is labeled with bounding boxes for visible tools, offering a challenging dataset for detecting and identifying surgical instruments in complex environments. The dataset provides video frames in JPEG format and corresponding annotations in CSV format, making it ideal for training and evaluating computer vision models in surgical tool detection.

The SOCAL dataset is particularly challenging due to the high degree of variability in instrument positioning, lighting conditions, and occlusions. Furthermore, the dataset includes instances where multiple instruments overlap, or partial views of instruments are obstructed by surgical hands, making detection a complex task. These factors provide an excellent test bed for evaluating the robustness of computer vision models in real-world surgical environments.

In previous studies, deep learning models such as RetinaNet and AutoML[9] have been explored for detecting surgical tools. RetinaNet[7], while achieving good precision, often struggles in real-time applications due to its relatively slower inference speed, especially on large, high-resolution datasets like SOCAL. Similarly, AutoML[8], which automates the process of model selection and optimization, has shown

*Conference acronym 'XX, June 03–05, 2018, Woodstock, NY*

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

promise but lacks the fine-tuning capabilities necessary for datasets with complex instrument occlusions and variations [2].

Some studies have also implemented transfer learning on well known object detection called YOLO (You look only once)[6]. YOLO model is really famous for fast real time object detection and with less computational power. It is also easy to code for transfer learning on custom dataset. One research [3] have implemented YOLOv4 on a similar dataset we have with a similar set of classes and they achieved a good mAP of 87.5.

Further improvements to the YOLO architecture, such as YOLOv7, have also been applied to the task of surgical instrument detection, achieving even higher accuracy. For instance, YOLOv7 has been shown to achieve a mAP of 95.8% in real-time instrument detection applications, making it one of the most promising models for this task [4]. Another study compared multiple models using IBM's Visual Analytics tool, reporting an accuracy of 91% for surgical instrument detection, further emphasizing the effectiveness of deep learning models in this domain [5].

In this study, we propose to evaluate and compare the performance of two state-of-the-art object detection architectures, Faster R-CNN and YOLO, for detecting neurosurgical instruments in the SOCAL dataset. Faster R-CNN, a two-stage detector, is known for its high precision, particularly in complex scenarios involving small or overlapping objects. However, its inference speed tends to be slower, which can be a limitation in real-time applications. On the other hand, YOLO (You Only Look Once), a single-stage detector, is designed for real-time object detection with faster processing speeds, but it may compromise accuracy when detecting smaller or partially occluded objects. By comparing these models on the SOCAL dataset, which presents challenges such as variable instrument positioning, occlusion, and lighting conditions, we aim to provide a comprehensive evaluation of the trade-offs between detection speed and accuracy. This comparison will offer insights into the suitability of each model for real-time surgical instrument detection, highlighting the advantages and limitations of both architectures in surgical environments.

## 2 Proposed Contribution

Our project build upon prior efforts to detect surgical instruments using advanced object detection techniques, specifically Faster R-CNN and YOLO to neurosurgical data. A significant enhancement in our approach involves the use of transfer learning to improve the performance[10]. As training a model from scratch could be computationally expensive and inefficient, we initialize our model with weights pre-trained on large scale object detection datasets such as COCO. This transfer learning approach allows the model to leverage the general object recognition knowledge gained from the datasets and fine-tune it for the specific task of neurosurgical instrument detection. Our primary contribution is a YOLO model that achieves higher mean Average Precision (mAP) scores than prior methods, improving both precision and recall, particularly for complex instruments. By transfer learning, we enhance the model's efficiency in detecting instruments with challenging environment of surgical procedures. This approach not only speeds up the detection process but also makes the model more scalable for real time surgical applications.

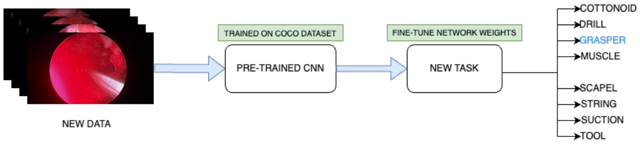


**Figure 1.** Transfer Learning Process for YOLO Model

## 3 Evaluation Plan

To assess the effectiveness of our model, we plan to take the following steps

### 3.1 Performance Metrics

We will evaluate our model using standard performance metrics in object detection, including:

- **Mean Average Precision (mAP)**: It is calculated by finding Average Precision(AP) for each class and then average over a number of classes.

$$\mathrm{mAP} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{AP}_n \tag{1}$$

- **Precision**: It measures how often our model correctly predict a target class.

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{2}$$

- **Recall**: It measures whether our model can find all the objects of the target class.

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{3}$$

- **F1-Score**: This is calculated as the harmonic mean of the precision and recall scores.

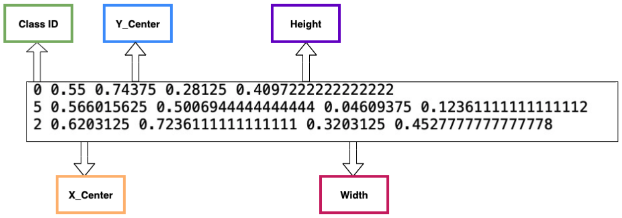
$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

### 3.2 Comparison with Baselines

Our model's performance will be compared against existing models like AutoML model, and RetinaNet.

## 4 Primary Experiments

As part of this study, we collected a dataset consisting of a folder of images totaling 3.2 GB in size. The corresponding annotations for each image are stored in the SOCAL.csv file, where each row contains the image path, along with the x, y coordinates, width, height parameters, and the class label in the final column. The dataset includes eight distinct surgical tools, with the following distribution: cottonoid (10,005 annotated instances), drill (210), grasper (15,943), muscle (4,560), scalpel (4), string (11,917), suction (22,356), and a general tool class (76). To facilitate model training, we have converted this dataset into a COCO-friendly format, where each image is associated with a corresponding text file that contains the class label and its x, y coordinates with the width and height of the bounding boxes. This conversion allows us to easily utilize state-of-the-art object detection models like Faster R-CNN and YOLO for further analysis.



```
0 0.55 0.74375 0.28125 0.4097222222222222
5 0.566015625 0.5006944444444444 0.04609375 0.12361111111111112
2 0.6203125 0.7236111111111111 0.3203125 0.4527777777777778
```

**Figure 2.** Annotated data in the YOLO format

## 5 Work Distribution

Shubham Laxmikant Deshmukh will be responsible for overseeing the transfer learning and training of the Faster R-CNN model(ResNet50, ResNet150) for detecting surgical instruments in real-time. Meanwhile, Pradyumna will focus on the transfer learning and training of the YOLO model for the same task. After the individual model training, we will collaboratively compare the performance of both models to determine which performs best on the SOCAL dataset. Based on these results, we will explore potential improvements and propose better solutions for future work, drawing from the insights gained through this comparison.

## References

[1] Kugener et al. "Utility of the Simulated Outcomes Following Carotid Artery Laceration Video Data Set for Machine Learning Applications." JAMA Network Open. 2022;5(3):e223177. doi:10.1001/jamanetworkopen.2022.3177

[2] Kugener G, Pangal DJ, Cardinal T, Collet C, Lechtholz-Zey E, Lasky S, Sundaram S, Markarian N, Zhu Y, Roshannai A, Sinha A, Han XY, Papyan V, Hung A, Anandkumar A, Wrobel B, Zada G, Donoho DA. Utility of the Simulated Outcomes Following Carotid Artery Laceration Video Data Set for Machine Learning Applications. JAMA Netw Open. 2022 Mar 1;5(3):e223177. doi: 10.1001/jamanetworkopen.2022.3177. PMID: 35311962; PMCID: PMC8938712.

[3] Y. Wang, Q. Sun, G. Sun, L. Gu and Z. Liu, "Object Detection of Surgical Instruments Based on YOLOv4," 2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM), Chongqing, China, 2021, pp. 578-581, doi: 10.1109/ICARM52023.2021.9536075.

[4] Zheng, L., Liu, Z. (2023). Real Time Surgical Instrument Object Detection Using YOLOv7. In: Stanimirović, P.S., Zhang, Y., Xiao, D., Cao, X. (eds) 6th EAI International Conference on Robotic Sensor Networks. ROSENET 2022. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi-org.ezproxy.lib.vt.edu/10.1007/978-3-031-33826-7

[5] Bamba, Y., Ogawa, S., Itabashi, M. et al. Object and anatomical feature recognition in surgical video images based on a convolutional neural network. Int J CARS 16, 2045–2054 (2021). https://doi-org.ezproxy.lib.vt.edu/10.1007/s11548-021-02434-w

[6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[7] T. -Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2999-3007, doi: 10.1109/ICCV.2017.324.

[8] Xin He, Kaiyong Zhao, Xiaowen Chu, AutoML: A survey of the state-of-the-art, Knowledge-Based Systems, Volume 212, 2021, 106622, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2020.106622.

[9] Unadkat V, Pangal DJ, Kugener G, Roshannai A, Chan J, Zhu Y, Markarian N, Zada G, Donoho DA. Code-free machine learning for object detection in surgical video: a benchmarking, feasibility, and cost study. Neurosurg Focus. 2022 Apr;52(4):E11. doi: 10.3171/2022.1.FOCUS21652. PMID: 35364576.

[10] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," in Proceedings of the IEEE, vol. 109, no. 1, pp. 43-76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.