

dx.doi.org/10.17488/RMIB.45.3.2

E-LOCATION ID: 1434

UMInSe: An Unsupervised Method for Segmentation and Detection of Surgical Instruments based on K-means

UMInSe: Método no Supervisado para la Segmentación y Detección de Instrumentos Quirúrgicos Basado en K-means

Rodrigo Eduardo Arevalo-Ancona¹ , Daniel Haro-Mendoza²  , Manuel Cedillo-Hernandez¹ , Victor J. Gonzalez-Villela² 

¹Instituto Politécnico Nacional, Ciudad de México - México

²Universidad Nacional Autónoma de México, Ciudad de México - México

ABSTRACT

Surgical instrument segmentation in images is crucial for improving precision and efficiency in surgery, but it currently relies on costly and labor-intensive manual annotations. An unsupervised approach is a promising solution to this challenge. This paper introduces a surgical instrument segmentation method using unsupervised machine learning, based on the K-means algorithm, to identify Regions of Interest (ROI) in images and create the image ground truth for neural network training. The Gamma correction adjusts image brightness and enhances the identification of areas containing surgical instruments. The K-means algorithm clusters similar pixels and detects ROIs despite changes in illumination, yielding an efficient segmentation despite variations in image illumination and obstructing objects. Therefore, the neural network generalizes the image features learning for instrument segmentation in different tasks. Experimental results using the JIGSAWS and EndoVis databases demonstrate the method's effectiveness and robustness, with a minimal error (0.0297) and high accuracy (0.9602). These results underscore the precision of surgical instrument detection and segmentation, which is crucial for automating instrument detection in surgical procedures without pre-labeled datasets. Furthermore, this technique could be applied in surgical applications such as surgeon skills assessment and robot motion planning, where precise instrument detection is indispensable.

KEYWORDS: JIGSAWS database, K-means, surgical instruments segmentation, unsupervised segmentation

RESUMEN

La segmentación de instrumentos quirúrgicos en imágenes es crucial para mejorar la precisión y eficiencia en cirugía, pero actualmente depende de anotaciones manuales costosas y laboriosas. Un enfoque no supervisado es una solución prometedora para este desafío. Este artículo introduce un método de segmentación de instrumentos quirúrgicos utilizando aprendizaje automático no supervisado, basado en el algoritmo K-means, para identificar Regiones de Interés (ROI) en imágenes y crear el *ground truth* de las imágenes para el entrenamiento de redes neuronales. La corrección Gamma ajusta el brillo de la imagen y mejora la identificación de áreas que contienen instrumentos quirúrgicos. El algoritmo K-means agrupa píxeles similares y detecta las ROI a pesar de los cambios en la iluminación, logrando una segmentación eficiente a pesar de las variaciones en la iluminación de la imagen y los objetos obstructores. Por lo tanto, la red neuronal generaliza el aprendizaje de las características de la imagen para la segmentación de instrumentos en diferentes tareas. Los resultados experimentales utilizando las bases de datos JIGSAWS y EndoVis demuestran la efectividad y robustez del método, con un error mínimo (0.0297) y alta precisión (0.9602). Estos resultados subrayan la precisión en la detección y segmentación de instrumentos quirúrgicos, lo cual es crucial para automatizar la detección de instrumentos en procedimientos quirúrgicos sin conjuntos de datos pre-etiquetados. Además, esta técnica podría aplicarse en aplicaciones quirúrgicas como la evaluación de habilidades del cirujano y la planificación de movimientos de robots, donde la detección precisa de instrumentos es indispensable.

PALABRAS CLAVE: base de datos JIGSAWS, K-means, segmentación instrumentos quirúrgicos, segmentación no supervisada

Corresponding author

TO: Daniel Haro-Mendoza

INSTITUTION: Universidad Nacional Autónoma de México

ADDRESS: Centro de Ingeniería Avanzada,
Departamento de Mecatrónica, Facultad de Ingeniería,
Universidad Nacional Autónoma de México, Coyoacán,
04510, Ciudad de México, México

EMAIL: danielharo@comunidad.unam.mx

Received:

7 May 2024

Accepted:

22 August 2024

INTRODUCTION

Minimally invasive surgery (MIS) represents a significant advancement in surgical procedures by reducing the complexity and enhancing the success rate of surgeries. These procedures improve surgeons' control over their instruments, leading to more precise operations. Additionally, MIS techniques significantly decrease patient recovery time and infection risks. The use of small incisions minimizes patient discomfort and faster healing. These advantages reduce wound exposure and its possible adverse effects, such as infections, and result in shorter hospital stays and quicker recoveries^[1].

A critical and indispensable requirement of MIS is the accurate detection and segmentation of surgical instruments. This process provides essential information about the location of the instruments, allowing for better planning of subsequent movements and reducing the chances of harming the patient. Detecting surgical instruments in images during surgery is crucial, as these procedures rely on real-time video data captured during the operation. Furthermore, the detection of surgical instruments has applications beyond the operating room. It can be used to practice surgical techniques, evaluate surgeons' skills, and perform detailed analyses for improved surgery planning. This approach enhances the precision and safety of surgical procedures and contributes to surgeons' ongoing education and training, ultimately leading to better patient outcomes.

The detection and segmentation of surgical instruments during MIS present several significant challenges. These challenges are related to different factors, including noise or image distortions caused by interference and fluctuations in illumination. Such issues produce low contrast between the surgical instruments and the surrounding tissues or background within the acquired images. These complexities highlight the necessity for advanced technological solutions to ensure accurate and reliable instrument detection.

Effective segmentation of surgical instruments is crucial for several reasons^{[2][3]}. Firstly, it produces a precise visualization of the tools used during surgery, essential for enhancing the safety, precision, and efficiency of the procedures^{[4][5]}. In addition, accurate visualization aids in surgical planning and execution, benefiting doctors and patients. Automating the segmentation process can simplify surgical workflows, enabling more focused surgical research and the development of advanced systems for skill evaluation and training. Automated segmentation systems play a pivotal role in improving the overall surgical process. This technique developed a sophisticated training and simulation environments where surgeons can improve their skills in a controlled, virtual setting before applying them in real-life situations. These systems enhance a surgeon's abilities, dexterity, and precision, contributing to better surgical outcomes^{[6][7]}. Moreover, integrating computer vision systems into surgical practices and training can significantly advance the field. These technologies enable the development of new, surgery-focused innovations that improve surgical techniques and patient care. Implementing these technologies in the medical field enhances the capabilities of surgical procedures, leading to more effective treatments and better patient outcomes^{[8][9]}.

In summary, the detection and segmentation of surgical instruments in minimally invasive surgeries are critical components that face several technical challenges. Addressing these challenges with advanced technological solutions can significantly improve surgical procedures' safety, precision, and efficiency. Automated segmentation and computer vision systems aid in real-time surgery and provide valuable tools for surgical training and skill development, paving the way for future advancements in medical technology and patient care.

This paper introduces a novel approach for the image segmentation of surgical instruments by implementing an unsupervised image segmentation algorithm. The primary objective of the proposed method is to accurately detect and segment suturing surgical instruments to determine their spatial location within an image. This method automates the generation of labels necessary for training a neural network, thereby addressing two critical challenges in the field. Firstly, the proposed approach eliminates the need for manual label generation. Instead, it utilizes the K-means algorithm to segment the image, identifying regions of interest (ROIs) where the suturing instruments are located. This automated label generation facilitates the training process for the neural network. Secondly, employing a neural network for image segmentation allows for the generalization of suturing instrument detection. This capability enables the trained model to identify and segment suturing instruments across various tasks and scenarios, not limited to a specific context. This versatility is demonstrated using different scenarios from the JIGSAW and EndoVis datasets. The proposed method offers an efficient and effective solution for the segmentation of surgical instruments in images. By automating label generation and leveraging the power of neural networks, this approach enhances the accuracy and applicability of surgical instrument detection and segmentation across diverse surgical scenarios. The main contributions of this approach are:

1. The K-means application for the detection of surgical instruments to create image labeling for the neural network training. This algorithm reduces errors by determining the similarity between pixels despite illumination changes, image distortions, noise adding, or video brightness changes.
2. The K-means algorithm, known for its efficiency, swiftly identifies patterns and segments images into homogeneous regions, providing a streamlined and effective image labeling process.
3. Automatically generated labels facilitate the neural network training process, improving the efficiency and effectiveness of the segmentation process.
4. Implementing a neural network-based segmentation approach that generalizes the detection of suturing instruments across various tasks and scenarios demonstrates versatility and robustness.
5. The proposed method is validated using different scenarios from the JIGSAW, Endoscape and EndoVis datasets, showcasing its applicability and effectiveness in multiple contexts.
6. Offering an efficient and effective solution for surgical instrument segmentation that enhances the accuracy and broad applicability of surgical instrument detection and segmentation across diverse surgical scenarios.

The rest of the paper is divided into sections: Section 2 describes a literature review of surgical instrument segmentation techniques to provide a better understanding. Section 3 describes materials and the proposed method. Section 4 presents the experimental results. Section 5 discusses the results obtained and analyzes why the proposed method outperforms other approaches. Finally, Section 6 concludes this paper.

Literature review

This section describes some techniques employed for surgical instrument segmentation, providing context for this paper's proposed method. Most recent and representative works for image surgical instrument segmentation reported in the scientific literature are depicted in Table 1. This literature review provides a better understanding of the current methods for segmenting surgical instruments into different tasks, which is the basis for this work.

TABLE 1. Literature review. (Continue in the next page).

Author	Description	Application	Efficiency	Database
Attia, <i>et al.</i> ^[10]	It utilizes recurrent neural networks and long short-term memory networks to determine relationships and learn dependencies between neighboring pixels.	Endoscopic images	Accuracy = 93.3 % IoU = 82.7 %	MICCAI 2016
Papp, <i>et al.</i> ^[11]	In this approach, different neural networks (UNet, TerausNet-11, TerausNet-16, Linknet-34) were used to create a general method for tool segmentation. The trained models were compared with their pre-trained counterparts, and the results show that the pre-trained models have lower accuracy than the trained ones.	Trained with endoscopic images and tested with suturing surgical instruments.	Accuracy = 97.3 % IoU = 70.96 % Dice = 79.91 %	Trained with MICCAI 2016 and tested with JIGSAW database.
Rahbar, <i>et al.</i> ^[12]	This approach employed an enhanced U-Net with the GridMask (EUGNet) data augmentation technique, designed to improve the performance of the proposed deep learning model.	Endoscopic images	Accuracy = 86.3 % IoU=80.6 % Dice = 89.5 %	da Vinci Research Kit (dVRK) open-source platform. Videos for testing our algorithm from open sources on the Internet, including the U.S. National Library of Medicine. The binary segmentation EndoVis 17 dataset.
Colleoni, <i>et al.</i> ^[13]	This approach combines robotic instrument simulation with artificial surgical images generated by a Cycle-GAN to train a U-Net model for surgical instrument segmentation.	Suturing surgical instruments	IoU = 86.3%	UCL Dataset, MICCAI '17 Dataset. Robot Assisted Radical Prostatectomy (RARP45) Dataset.
Deepika, <i>et al.</i> ^[14]	A pretrained region-based convolutional neural network (R-CNN) model was used. The original classification head was replaced with a new layer consisting of 6 outputs. This modified network was subsequently fine-tuned using our annotated neurosurgical video dataset to enhance its performance for the specific task of surgical instrument detection.	Neurosurgery	IoU = 96% Precision = 96.7%	The dataset consists primarily of 5 instruments which are commonly used in neurosurgery such as Suction, Bipolar Forceps, Straight Needle Holder, Straight Micro Scissor and Dural Tooth Forceps.
Leifman, <i>et al.</i> ^[15]	Integrate synthetic images into the training workflow with the help of a CycleGAN. Using a dataset of laparoscopic images paired with their bounding box annotations, we automatically produce pixel-perfect segmentations through the application of DeepMAC. This technique enhances instance segmentation by leveraging CenterNet.	Laparoscopic instruments	Dice = 89% Accuracy = 93%	Endoscopic Vision 2015 Instrument Segmentation and Tracking Dataset. EndoVis2019.

TABLE 1. Literature review. (Continue in the next page).

Author	Description	Application	Efficiency	Database
Mishra, <i>et al.</i> ^[16]	A transfer learning is used in the neural network to extract the background and foreground of the image for endoscopic instruments segmentation.	Endoscopic instruments	Accuracy = 89%	Not mentioned.
Lou, <i>et al.</i> ^[17]	A Min-Max Similarity (MMS) approach utilizes a contrastive learning framework for dual-view training by employing classifiers and projectors.	Endoscopic instruments	Dice = 93% IoU = 89%	EndoVis 17. ART-NET. RoboTool.
Colleoni and Stoyanov ^[18]	Used train deep learning models cycle-GAN and MUNIT frameworks using image-to-image translation techniques.	Endoscopic images	IoU = 96%	MICCAI 2015 EndoVis
Jha, <i>et al.</i> ^[19]	A Dual decoder attention network (DDANet) is implemented.	Laparoscopic surgeries	Dice = 87% IoU = 81% Recall = 87% Precision = 93 % Accuracy = 98 %	Robust Medical Instrument Segmentation
Allan, <i>et al.</i> ^[20]	Perform a position estimation of surgical instruments. Surgical instruments are segmented them to identify their location using silhouette detection and optical flow.	Laparoscopic surgeries	Precision = 87 % Recall = 93 % F1 score = 90 %	Da Vinci LND dataset
Wang, <i>et al.</i> ^[21]	Clustering similar pixels using the random forest algorithm. Subsequently, they perform the 3D position estimation of the surgical instruments by calculating their kinematics.	Endoscopic images	IoU = 82 %	Not mentioned
Yu, <i>et al.</i> ^[22]	Segment the surgical instruments using convolutional neural networks. The neural network used for the segmentation task is based on the U-Net model, which they modified to obtain a feature map that eliminates the need to crop the image, focusing on the specific area where the surgical instruments are located.	Endoscopic images	Accuracy = 91 % IoU = 86 % Dice = 92 %	Robotic Instrument Segmentation Challenge.

TABLE 1. Literature review. (Continue from previous page).

Author	Description	Application	Efficiency	Database
Xue and Gu ^[23]	Implemented a surgical instrument segmentation method based on the MobileNetV2 neural network model, in which they added the Atrous Spatial Pyramid Pooling layer to focus on spatial features of the image using the Convolutional Block Attention Module CBAM to improve the efficiency of the neural model.	Common instruments	IoU = 86 % Accuracy = 88 %	No public dataset, the dataset consists in 7 common surgical instruments with a data augmentation.
Baby, <i>et al.</i> ^[24]	MaskRCNN model for surgical instrument segmentation. In addition, they added a layer for their classification (bipolar forceps, prograsp forceps, large needle driver, vessel sealer/suction Instrument, grasping retractor/ clip applicator, monopolar curved scissors, ultrasound probe).	Endoscopic images	IoU = 72 %	Robotic Instrument Segmentation Challenge.
Streckert, <i>et al.</i> ^[25]	The images were created by placing surgical instruments on a green screen where surgical procedures replaced the background. Additionally, to increase the dataset size, they used a GAN to generate more images of surgical procedures. Subsequently, two neural models based on the SegNet network were trained for feature learning.	Endoscopic images	Dataset Endovio: IoU = 91 % Synthetic dataset: IoU = 89 %	Robotic Instrument Segmentation Challenge. ^[28]
Yamada, <i>et al.</i> ^[26]	Employs hierarchical clustering to automatically detect key events and changes in the surgical workflow.	Surgeons practice	Accuracy = 88 %	Not mentioned
Zhang, <i>et al.</i> ^[27]	Employment of surgical tools with detailed textures as annotation samples and a WGAN-GP.	Endoscopic surgery	IoU = 92 %	Dataset 1 and 2 are recorded using the STRAS robot
Qayyum, <i>et al.</i> ^[28]	The region of interest is cropped to reduce processing time and the U-Net model is applied for image segmentation.	Endoscopic	Accuracy = 95 % F1 score = 95 %	MICCAI 2022

MATERIALS AND METHODS

In this chapter, we outline the datasets, tools, and methodologies employed in the development and validation of our proposed algorithm for the segmentation of surgical instruments. The approach combines advanced image processing techniques, unsupervised learning algorithms, and neural network training to achieve accurate and efficient segmentation. We first describe the datasets used in the experiments, including the JIGSAWS, MICCAI 2015 EndoVis and Endoscope databases, which provide the necessary data for training and testing. Next, we detail the preprocessing steps applied to the images, including brightness adjustment and binarization, which are crucial for enhancing the visibility and differentiation of surgical instruments. Finally, we present the segmentation method, focusing on the stages of approximate detection, region of interest identification, and neural network training.

Description of the databases used

In this section, we provide an overview of the databases utilized for developing and validating our segmentation algorithm. We discuss the JIGSAWS database and the MICCAI 2015 EndoVis database, highlighting their relevance and features.

The **JIGSAWS (JHU-ISI Gesture and Skill Assessment Working Set) database**^[29] evaluates surgical skills in minimally invasive surgery procedures at Johns Hopkins University. The presented algorithm training and validation used the experimental data from the freely available JIGSAWS database, which contains stereoscopic video data and kinematic data of the position and orientation of the tips of the da Vinci system forceps. Both data sets were captured by the da Vinci™ Robot API with a sampling rate of 30Hz. The test subjects included eight surgeons with different skill levels in robotic teleoperation: two expert surgeons with over 100 hours of experience, two intermediate surgeons with 10-100 hours of experience, and four novice surgeons with less than 10 hours of experience. The experiments requested from the surgeons involved performing surgical tasks such as suturing, needle passing, and knot tying. The videos are annotated with specific gestures, skill assessments, and sensor data from the da Vinci robotic system, providing detailed analysis of movements and techniques. Used to develop machine learning algorithms, this database facilitates automated skill assessment, surgical training with real-time feedback, and research to improve surgical techniques. Each video contains a total of 1794 frames, with a frame width of 640 pixels and a frame height of 480 pixels.

The **MICCAI 2015 EndoVis database**^[30] is a set of high-resolution videos recorded during endoscopic surgical procedures. It was created to drive the development of advanced algorithms for detecting, segmenting, and tracking surgical instruments. Each video is annotated with precise labels, segmentation masks, and instrument motion trajectories, facilitating research in computer vision applied to surgery. This database is an essential tool for improving the precision and effectiveness of surgical procedures. This database contains 160 images with a size of 640x480.

The **Endoscape 2023 database**^[31] This dataset presents laparoscopic videos designed for surgical anatomy and tool segmentation, object detection, and the assessment of the Critical View of Safety during procedures. The dataset offers a comprehensive collection of annotated videos, providing a robust foundation for research and development in surgical image analysis.

Description of the proposed method

This section outlines our segmentation method for surgical instruments. Figure 1 illustrates the comprehensive workflow of our proposed method, offering a detailed visual representation of the segmentation process. This process employs the K-means algorithm for the masks or labels generation by clustering pixels. This precise segmentation of surgical instruments enables their subsequent detection and localization. The real potential lies in the neural network. The neural network is trained to generalize the learning of the surgical instrument's features. The trained model identifies the surgical instruments in videos related to different tasks, providing new possibilities in surgical technology, inspiring a future where surgical procedures are more accurate and efficient. This approach leverages advanced self-learning algorithms, ensuring accurate segmentation of surgical instruments for developing surgical guidance and automation systems. These advancements aim to improve surgical outcomes and patient safety by providing greater accuracy and efficiency in identifying and manipulating surgical instruments during procedures.

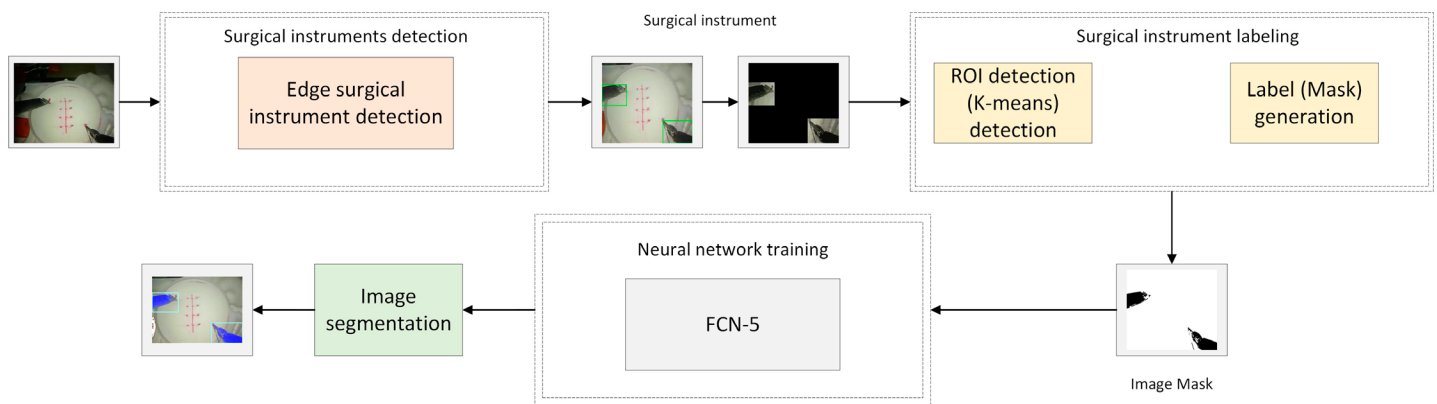


FIGURE 1. Unsupervised surgical instruments segmentation general diagram.

The proposed method for the segmentation of surgical instruments consists of three detailed stages:

1. **Approximate Detection of Surgical Instruments:** This initial stage aims to reduce computational costs by focusing on specific image areas. This task is achieved through edge detection techniques, which identify and outline the boundaries of potential surgical instruments. Concentrating computational resources on these edges effectively narrows down the regions of interest, streamlining the subsequent segmentation process.
2. **Identification of Regions of Interest for image labeling:** In this stage, the focus is on the areas identified in the previous step. The K-means clustering algorithm segments the image by grouping similar pixels. This segmentation process identifies specific areas where the surgical instruments are located. Once these regions are located, the image is binarized to create clear distinctions between the instruments and the background. This binarized image generates masks or labels essential for training the neural network.
3. **Neural Network Training:** The final stage involves training a neural network using the masks automatically generated in the previous step. These masks serve as labels, providing the neural network with accurate examples of surgical instruments. The training focuses on generalizing feature learning, enabling the neural network to detect and segment surgical instruments across various tasks and scenarios. This generalization ensures the neural network can perform robustly in different surgical environments, enhancing its applicability and reliability.

This method ensures efficient, accurate, and adaptable segmentation of surgical instruments through edge detection for initial focus, K-means clustering for precise segmentation, and neural network training for generalized detection. This comprehensive approach lays the groundwork for advanced surgical guidance systems, improving surgical outcomes and patient safety through enhanced instrument detection and segmentation capabilities.

Approximate detection of surgical instruments

In this section, we describe the initial stage of our segmentation method for surgical instruments, focusing on reducing computational costs by narrowing down the areas of interest within the image. Figure 2 illustrates this stage, where edge detection techniques are employed to identify and outline the boundaries of potential surgical instruments. By concentrating computational resources on these edges, we streamline the segmentation process.

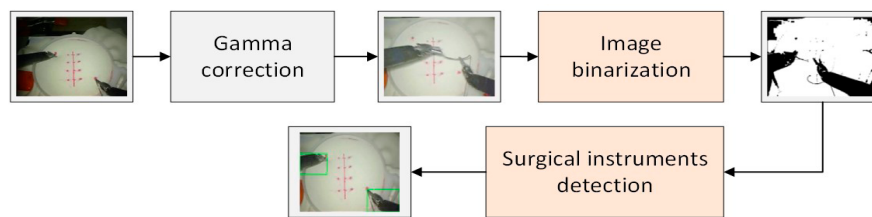


FIGURE 2. Image frame preprocessing diagram.

The image quality is modified by increasing the brightness to enhance surgical instrument detection using the Gamma correction^{[32][33]}. This adjustment is crucial for improving the precision in identifying preliminary Regions of Interest where surgical instruments are located. By brightening the image, the contrast between the surgical instruments and the surrounding tissues is increased, making the instruments more distinguishable. This step is essential for accurately focusing the detection process on the relevant areas of the image. Figure 3 illustrates this image preprocessing step, showing the difference in clarity and instrument visibility before and after the brightness adjustment.

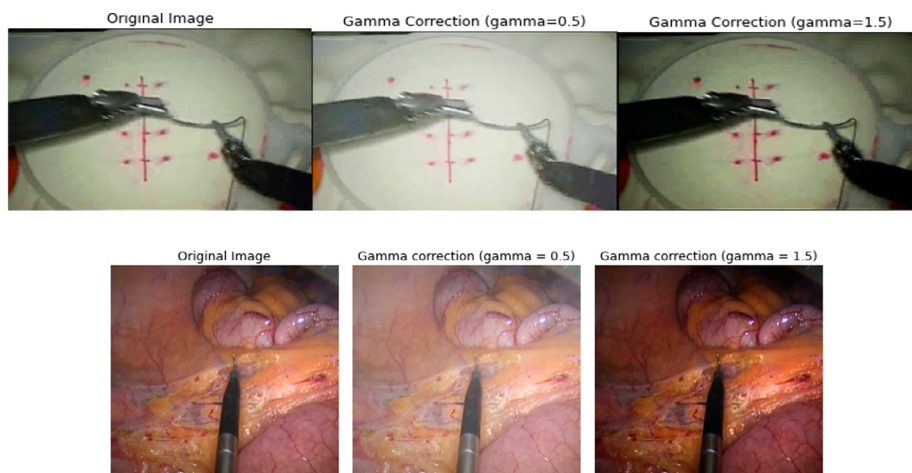


FIGURE 3. Image bright adjusts with different gamma values.

Figure 3 demonstrates the effects of with different γ values on image brightness. When γ is less than one, the image becomes brighter, enhancing the visibility of surgical instruments. Conversely, a γ value greater than one darkens the

image, which can be useful in different lighting conditions. This technique provides precise control over image luminosity, allowing for tailored adjustments that enhance the detection process. After gamma correction, the image is binarized to approximate the location of surgical instruments. Binarization converts the processed image into a binary format (black and white), simplifying the detection process. To automate the binarization step, the Otsu algorithm is employed^[34]. This algorithm analyzes the grayscale histogram of the image to determine the optimal threshold value that maximizes the variance between the foreground (surgical instruments) and the background. By doing so, it enhances the separation of the instruments from the surrounding tissues, ensuring accurate detection. Figure 4 showcases how different γ values affect image brightness and the subsequent steps of binarization and thresholding, culminating in a more efficient and precise detection of surgical instruments. This method leverages the combined power of gamma correction and optimal thresholding to improve the overall effectiveness of the surgical instrument detection process.

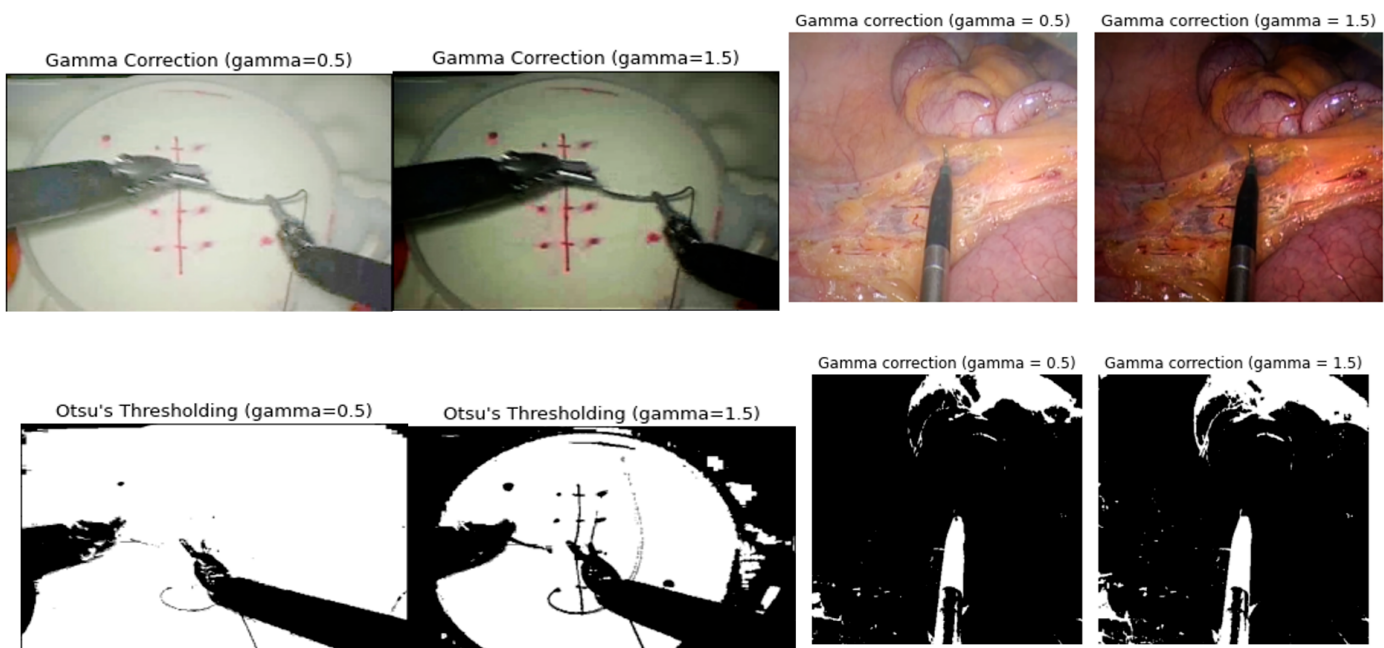


FIGURE 4. Image binarization with different image gamma correction.

Figure 4 illustrates the approximate location of surgical instruments under different brightness adjustments achieved through gamma correction. A gamma value less than one has been found to significantly enhance the algorithm's performance, leading to more efficient and accurate localization of surgical instruments.

Identification of regions of interest for image labeling

This section describes the use of unsupervised learning techniques for image segmentation, focusing on the identification of Regions of Interest (ROIs) in surgical images. It explains how the K-means algorithm is applied to cluster pixels with similar characteristics, such as color, texture, or intensity, enabling the precise detection and segmentation of surgical instruments within the images. This approach enhances segmentation accuracy, reducing errors and optimizing the generation of image labels for subsequent neural network training.

Machine learning algorithms are crucial in data analysis by constructing models that discern patterns and aid in decision-making processes^[35]. These algorithms are categorized into supervised, semi-supervised, and unsupervised learning

methods. Supervised learning utilizes labeled data to identify pattern-related features, while semi-supervised learning augments training efficiency by generating new data based on existing labels. In contrast, unsupervised learning, such as data clustering techniques, identifies patterns in data without relying on predefined labels. By analyzing and clustering data based on similarities or differences, these methods unveil diverse patterns within datasets, offering insights and uncovering hidden relationships crucial for tasks like image segmentation^[36].

Unsupervised image segmentation identifies and analyzes patterns and features within image regions^[37]. On the one hand, unsupervised methods do not require any label for pattern recognition, such as, autosupervised neural network models, which used the labels that are automatically generated by an algorithm like in this proposal. On the other hand, some algorithms cluster similar pixels together based on shared characteristics like color, texture, or intensity, enabling the identification of areas requiring detailed scrutiny due to their significance within the image. In surgical instrument segmentation, Regions of Interest contain information regarding the precise locations of instruments. Utilizing unsupervised learning techniques facilitates the detection of features and patterns autonomously, enhancing the accuracy of image analysis.

This paper used the K-means (Figure 5, Figure 6) algorithm to identify ROIs and segment surgical instruments effectively within images to generate image labels for the neural network training. By iteratively clustering similar pixels, K-means optimizes the detection of specific image areas, thereby reducing false positive errors in segmentation. Each pixel is assigned to a class based on its proximity to centroids, which represent characteristic points in the feature space of each class. This assignment is determined by minimizing the Euclidean distance between pixels and centroids, iteratively adjusted until convergence^{[38][39][40]}. The utilization of K-means enhances the precision of surgical instrument detection and segmentation (Table 2), demonstrating its efficacy in improving image analysis methodologies for surgical tasks.

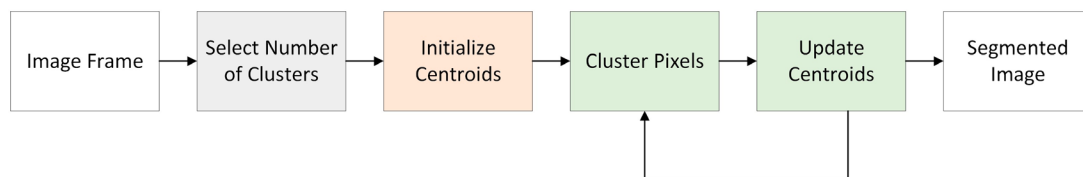




FIGURE 5. K-means flowchart.

Algorithm of image labeling generation:

1. Frame extraction ($I(x,y)$)
2. Gamma correction ($I'(x,y)$)
3. Aproximate surgical instrument detection
 - 3.1. Otsu thresholding
 - If $I'(x,y) < 0$
 $I_b(x,y) = 0$
 - Else
 $I_b(x,y) = 255$
4. Contour detection for surgical instrument detection
5. K-means segmentation
 - Number of cluster selection
 - Pixel clustering
6. Ground truth selection

FIGURE 6. Algorithm description for image labeling generation.

TABLE 2. Image ground truth generation.

Image Frame	Generated Segmentation Mask	Ground Truth
		

In conclusion, the precise identification of Regions of Interest through unsupervised learning techniques, such as the K-means algorithm, lays a solid foundation for effective surgical instrument segmentation. This step is crucial in generating accurate labels that are then used to train neural networks, ensuring that the system can generalize and reliably identify surgical instruments across a variety of scenarios.

Neural network training for surgical instrument generalize segmentation

In this section, we delve into the process of training a neural network to achieve generalized segmentation of surgical instruments across various datasets. The objective is to develop a robust model that, once trained on a specific dataset like JIGSAWS, can accurately segment surgical instruments in different datasets without the need for retraining. Utilizing a Fully Convolutional Network (FCN-5) architecture, this approach leverages automatically generated training labels and optimized hyperparameters to enhance segmentation accuracy. The following text will explore the details of the training process, the effectiveness of the chosen architecture, and the overall impact on the efficiency and reliability of surgical instrument segmentation.

The neural network is used to segment surgical instruments in images. The primary goal is to achieve effective generalization, meaning that once trained on the JIGSAWS dataset, the neural network can accurately segment surgical instruments in other data sets, regardless of variations in the specific instruments present in those data sets. For this purpose, a neural network architecture based on Fully Convolutional Networks, such as the FCN-5 model is used (Figure 7).

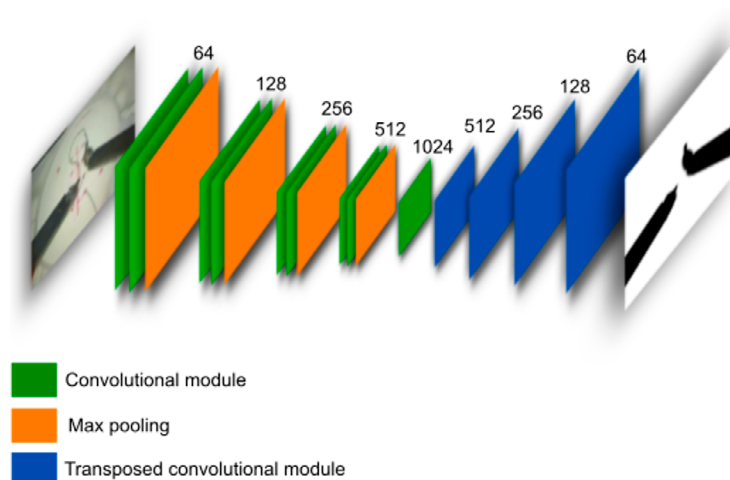


FIGURE 7. Fully Convolutional Networks (FCN-5) architecture.

TABLE 3. Training hyperparameters

Hyperparameters	
Optimizer	Adam
Learning Rate	0.00009
Epochs	30

The FCN-5 architecture is used for these types of applications since this neural network learns specific image features for surgical instrument segmentation. Table 3 describes the neural network's training hyperparameters, including details such as learning rate, number of training epochs, and other specific settings that influence the network's performance and generalization ability. These hyperparameters are essential to optimizing the training process and ensuring that the network can effectively handle the diversity of surgical instrument images. The training process involves using the JIGSAWS dataset to train the neural network. The proposed method automatically generates the necessary labels for training, eliminating the need for manual labeling. This is achieved through the use of the K-means algorithm and image binarization. The loss function used in this method is the Jaccard index, also known as Intersection over Union (IoU) (Equation 1). This metric evaluates the overlap between the model prediction and the ground truth. The IoU is calculated by dividing the intersection area between the prediction and the ground truth by the area of their union. A higher IoU value indicates higher segmentation accuracy, meaning the model prediction matches the ground truth better.

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (1)$$

The Intersection Area is the number of pixels that match the prediction and the ground truth. The Union Area, another important factor in the Jaccard index calculation, is the total number of pixels present in both the model prediction and the ground truth. It counts shared pixels only once, which is a key aspect in the accuracy evaluation of the segmentation.

After training, the neural network can generalize, applying the acquired knowledge to segment surgical instruments across different datasets accurately. This capability is achieved without retraining the neural network for each new dataset, showcasing the method's effectiveness and efficiency in diverse conditions.

A key contribution of this method is its ability to automatically generate labels for training the neural network. This automation reduces the time of image labeling. Additionally, the training process significantly enhances segmentation accuracy. This approach reduces the time required by eliminating the labor-intensive task of creating manual labels.

Moreover, the automated label generation process ensures that the neural network is exposed to various training examples, further improving its ability to generalize across different datasets. This robustness allows the network to effectively handle variations in surgical images, such as differences in light-

ing, instrument types, and surgical environments. The primary advantages of this method include:

1. The trained neural network can effectively segment surgical instruments in new datasets without retraining.
2. This approach reduces the need for manual intervention, accelerates the training process, and enhances accuracy, making it a time-saving boon for busy professionals in the field.
3. The approach ensures consistent and reliable results across different conditions and datasets

RESULTS AND DISCUSSION

In this section, we present a detailed evaluation of a surgical instrument segmentation algorithm, implemented using Python and tested on diverse datasets, including JIGSAWS and EndoVis. We delve into the algorithm's performance by comparing the generated segmentation masks with manually created ground truth, utilizing key metrics such as the Jaccard index, accuracy, precision, recall, F1 score, and Mean Squared Error (MSE). The impact of varying gamma values and cluster numbers on segmentation efficiency is analyzed, and the proposed method is contrasted with existing segmentation techniques in the literature. Furthermore, we assess the performance of different neural networks for surgical instrument segmentation, balancing precision and processing time to determine the most effective approach for surgical applications.


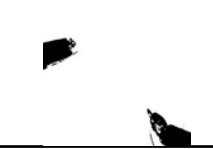
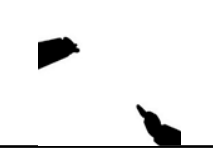
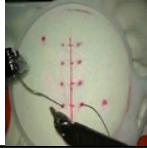

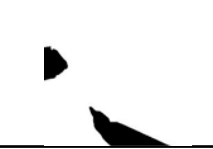


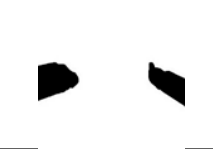
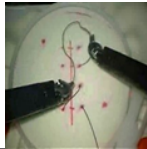
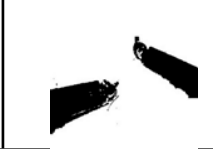
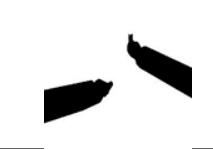

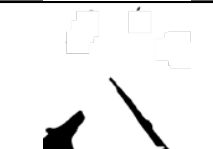
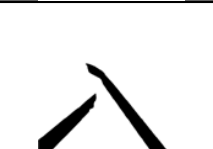
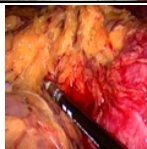
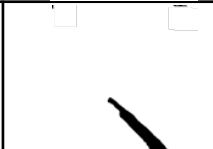
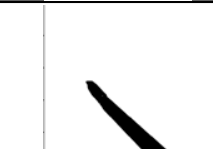



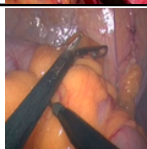


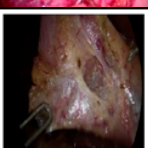





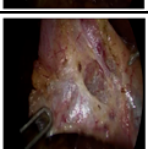


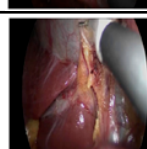


The algorithm outlined in this paper is implemented using the Python framework and executed on hardware equipped with an Intel(R) Core (TM) i7-6700 CPU @ 3.40GHz, 8 GB of RAM, and an NVIDIA GeForce GTX 960 GPU. The experiments were conducted using the Suring videos from the JIGSAWS dataset, Endoscape 2023 dataset and the **MICCAI 2015 EndoVis database**, which contained different videos and image frames.

The proposed system was assessed by comparing the generated mask in the surgical instrument segmentation and manually generated ground truth. This process involved detecting and segmenting surgical instruments to analyze their accuracy and effectiveness in image localization. This assessment validated the system's performance for surgical instrument segmentation.

Table 4 compares the generated mask through the segmentation proposed method to the created ground truth. The mask derived from unsupervised segmentation demonstrates a high similarity to the ground truth, providing precise surgical instrument image localization through the segmentation and detection process. The applications of the proposed algorithm encompass different tasks, such as assessing surgeons' skills by estimating instrument positions, establishing parameters to alert surgeons of potential instrument-organ proximity or collision risks, exploring novel surgical techniques, predicting trajectories, and enhancing risk mitigation during surgical procedures.

In order to determine the optimal value of gamma and optimize the number of clusters for efficient segmentation, a quantitative set of metrics to evaluate the system's performance are employed. These metrics include the Jaccard index (Intersection over Union), accuracy, precision, F1 score, Mean Squared Error (MSE), and Recall.

TABLE 4. Comparison among original JIGSAWS's image frame, generated segmentation mask and ground truth.

Image Frame	Generated Segmentation Mask	Ground Truth	Image Frame	Generated Segmentation Mask	Ground Truth
					
					
					
					
					
					

The Jaccard index (Equation 1) is a widely used metric for evaluating the performance of image segmentation and object detection algorithms. It quantifies the overlap detected pixels between the predicted segmentation and the ground truth, with values ranging from 0 to 1. A Jaccard index of 0 indicates no overlap between the regions, while a value of 1 signifies complete overlap. Thus, a higher Jaccard index indicates a better alignment between the predicted and ground truth regions. This metric is particularly valuable in image segmentation tasks, where the primary objective is to assess the accuracy of the predicted regions relative to the actual regions.

The Accuracy is used in classification problems to evaluate the performance of a model. It is defined as the proportion of correct predictions about the total predictions made by the model (Equation 2). Accuracy measures how well a model correctly classifies instances.

$$acc = \frac{\text{Correct pixels prediction}}{\text{Total predictions}} \quad (2)$$

Precision evaluates the pixels segmentation related to the ROI and it is compared to the ground truth. The image segmentation precision can be defined as the proportion of pixels correctly classified as part of the region of interest (true positives) relative to the total pixels classified as part of that region (true positives plus false positives) Equation 3.

$$precision = \frac{Tp}{Tp + Fp} \quad (3)$$

where true positives (TP) are the pixels correctly classified and false positives (FP) are pixels incorrectly classified. Recall measures the correct pixels of the class of interest concerning the total pixels of that class (Equation 4).

$$recall = \frac{Tp}{Tp + Fn} \quad (4)$$

where false negatives (FN) are the relevant pixels, the model has not identified as part of the region of interest. A high recall indicates that the model can effectively identify the most relevant pixels. The F1 Score evaluates the segmentation performance (Equation 5). It provides a balance between precision and recall.

$$F1 = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (5)$$

The Mean Squared Error (MSE) compares the ground truth with the values obtained by segmentation and measures the errors (Equation 6).

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \quad (6)$$

where y_i is the ground truth pixel value and \hat{y}_i is the predicted value.

TABLE 5. Surgical instrument segmentation efficiency applying different γ and clusters values.

	$\gamma = 0.5$					$\gamma = 1.5$					
	N = 5	N = 15	N = 25	N = 50	N = 150		N = 5	N = 15	N = 25	N = 50	N = 150
IoU	0.964	0.965	0.965	0.910	0.910	IoU	0.962	0.795	0.798	0.943	0.943
Acc	0.968	0.970	0.970	0.915	0.915	Acc	0.794	0.824	0.824	0.950	0.951
MSE	0.031	0.002	0.029	0.084	0.084	MSE	0.205	0.178	0.175	0.049	0.048
F1 score	0.933	0.935	0.935	0.746	0.746	F1 score	0.721	0.749	0.751	0.902	0.902
Preci- sion	0.937	0.939	0.939	0.909	0.909	Preci- sion	0.710	0.730	0.731	0.886	0.887
Recall	0.935	0.936	0.936	0.706	0.706	Recall	0.869	0.886	0.887	0.924	0.925

Table 5 and Figure 8 presents the system's surgical instruments segmentation performance with different gamma values, which adjust the image brightness. These experiments indicate that increasing image brightness improves the efficiency of segmentation and detection of surgical instruments. Furthermore, the image segmentation with different number of clusters is presented where the segmentation process is improved; however, this increase in efficiency comes with an increase in processing time. It is important to note that if the value of γ exceeds 1, the efficiency of image segmentation decreases. This effect is because a gamma value greater than one makes the image darker, making accurate identification of surgical instruments complex due to lower visibility and contrast in the resulting images.

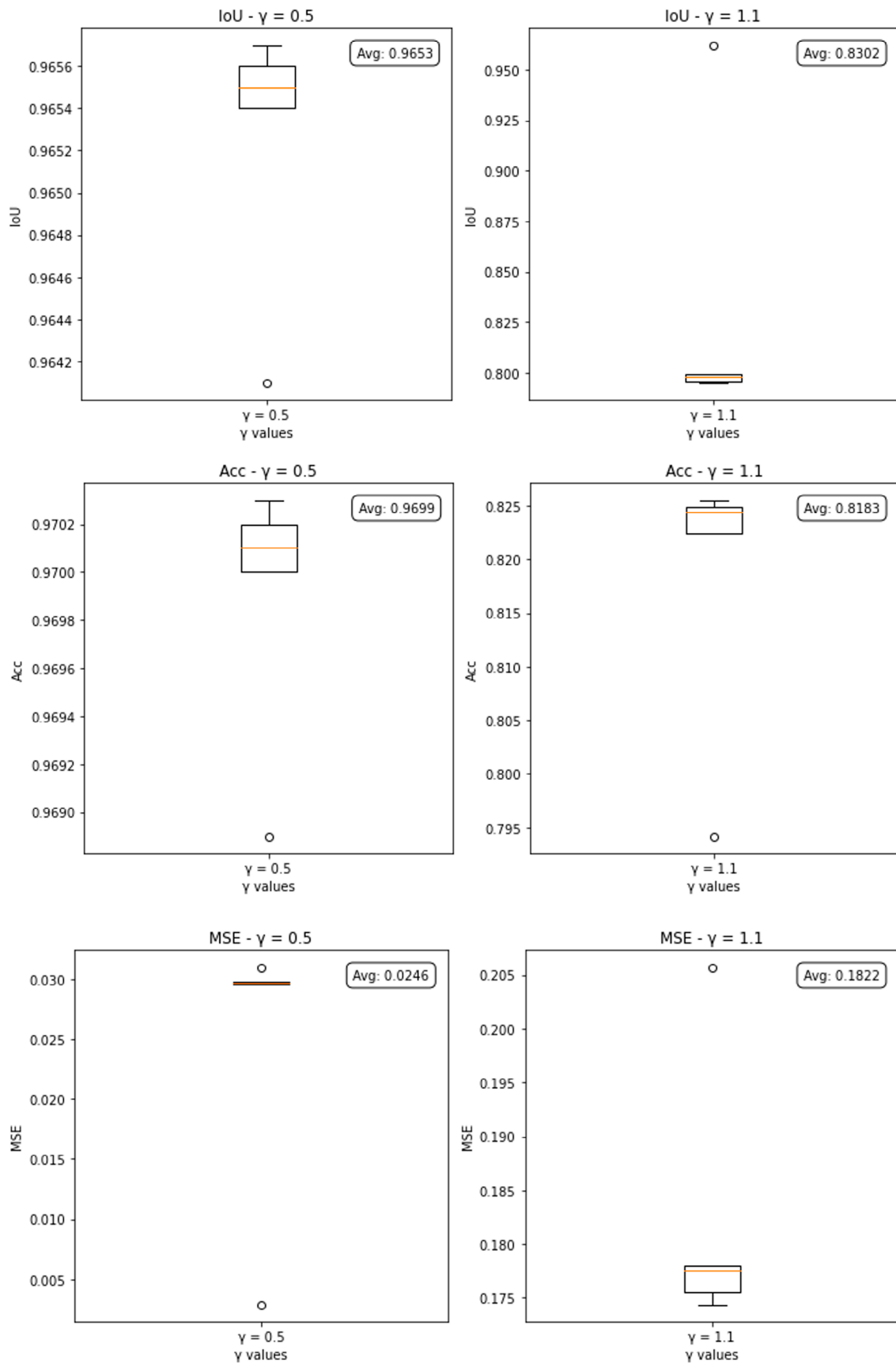


FIGURE 8. Surgical instruments segmentation efficiency. (Continue in the next page).

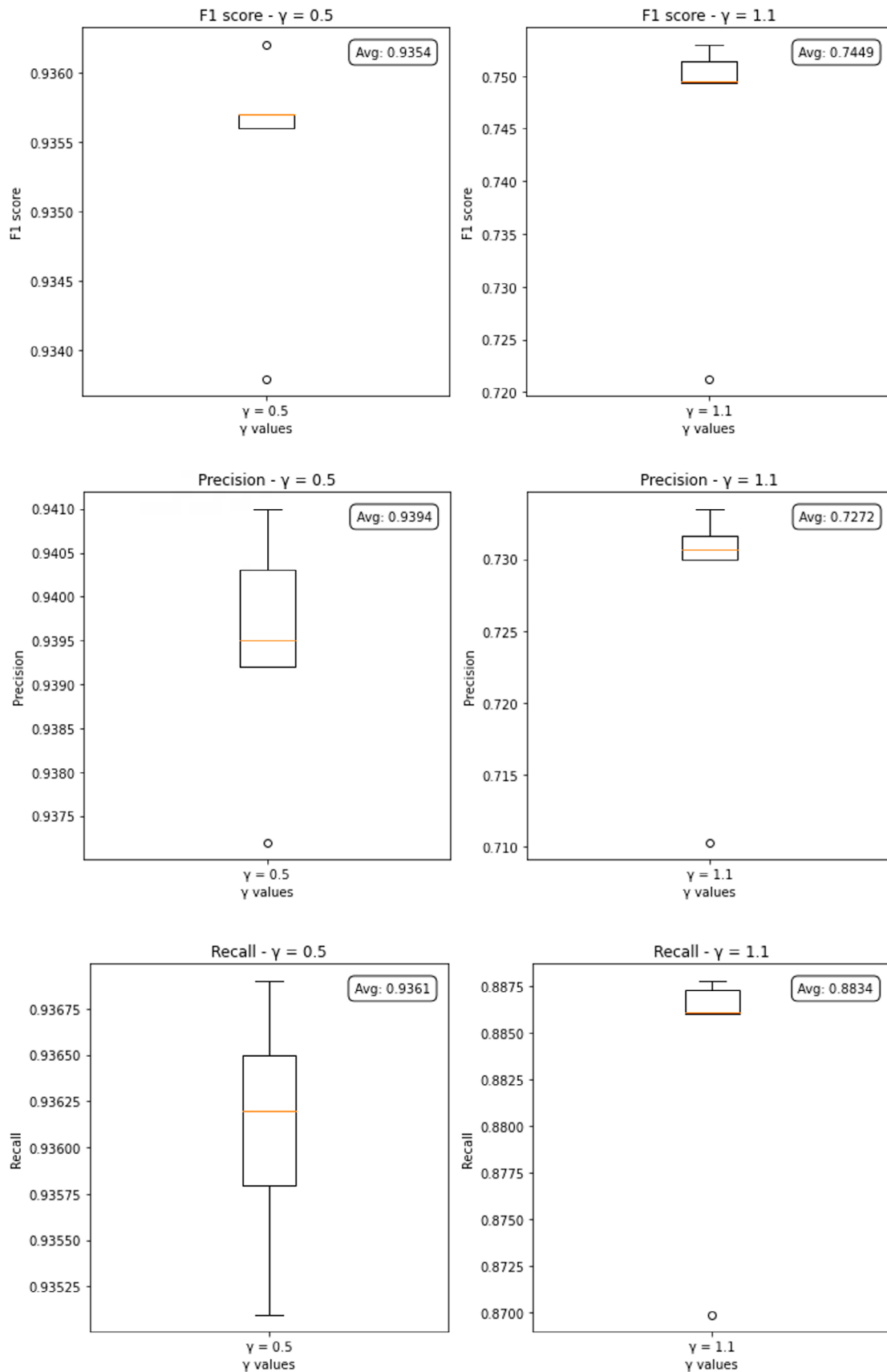


FIGURE 8. Surgical instruments segmentation efficiency. (Continue from previous page).

The results indicate that brightness is a key factor in image segmentation. A lower γ , more robust and accurate segmentation is achieved when a moderate number of clusters is used ($N \leq 25$). Specifically, the IoU and accuracy remain high up to several clusters of $N = 25$, decreasing from $N = 50$. The decrease in efficiency is because several regions of interest are generated, generating errors when binarizing the image and leading to the segmentation of surgical instruments where the error increases and the processing time increases. On the other hand, with a higher γ , a different behavior is observed. The initial precision is low with a small number of clusters ($N \leq 25$), but it improves significantly as the number of clusters increases, reaching optimal values from $N = 50$ because the algorithm allows for better detection of image details by dividing it into regions of interest.

Figure 9 demonstrates a correlation between the processing time and the number of clusters employed. The processing time increases if the number of clusters for image segmentation increases. Hence, it is necessary to balance processing time and desired efficiency. By optimizing this balance, we can ensure the system's efficient performance, delivering accurate results without sacrificing time efficiency. Upon analyzing the results with different gamma values and the number of clusters from extracted frames, we determined that the optimal gamma value is 0.5, coupled with a cluster count of 5 for pixel clusters. This configuration effectively provides a balance between precise surgical instrument identification, processing time, and segmentation efficiency, ensuring optimal system performance.

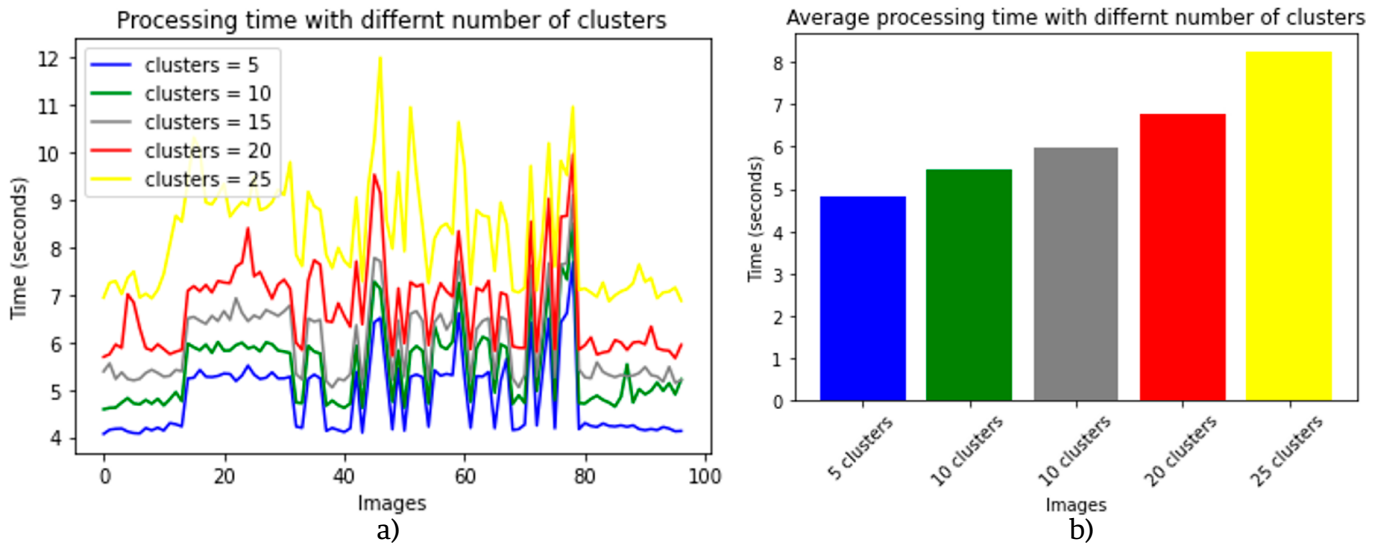


FIGURE 9. Processing time with different number of clusters.

TABLE 6. JIGSAWS instrument segmentation for tampered images.

Dataset	Attack	IoU	Attack	IoU	Attack	IoU
JIGSAWS	Salt and pepper noise adding 0.001	0.8570	Salt and pepper noise adding 0.005	0.8550	Salt and pepper noise adding 0.05	0.7963
	JPEG compression quality factor = 90	0.8528	JPEG compression quality factor = 70	0.7792	JPEG compression quality factor = 50	0.7524
	Median filter kernel = 3 x 3	0.9592	Gaussian filter kernel = 3 x 3	0.9581	Blurring kernel = 3 x 3	0.9593
	Median filter kernel = 7 x 7	0.9435	Gaussian filter kernel = 7 x 7	0.9546	Blurring kernel = 7 x 7	0.9498

TABLE 7. Endovis dataset instrument segmentation for tampered images.

Dataset	Attack	IoU	Attack	IoU	Attack	IoU
Endovis dataset	Salt and pepper noise adding 0.001	0.8537	Salt and pepper noise adding 0.005	0.8535	Salt and pepper noise adding 0.05	0.8403
	JPEG compression quality factor = 90	0.7898	JPEG compression quality factor = 70	0.7772	JPEG compression quality factor = 50	0.7504
	Median filter kernel = 3 x 3	0.7532	Gaussian filter kernel = 3 x 3	0.8062	Blurring kernel = 3 x 3	0.8021
	Median filter kernel = 7 x 7	0.7435	Gaussian filter kernel = 7 x 7	0.8007	Blurring kernel = 7 x 7	0.7954

TABLE 8. Endoscape dataset instrument segmentation for tampered images.

Dataset	Attack	IoU	Attack	IoU	Attack	IoU
Endoscape 2023	Salt and pepper noise adding 0.001	0.8510	Salt and pepper noise adding 0.005	0.8657	Salt and pepper noise adding 0.05	0.8653
	JPEG compression quality factor = 90	0.8718	JPEG compression quality factor = 70	0.8185	JPEG compression quality factor = 50	0.7653
	Median filter kernel = 3 x 3	0.8658	Gaussian filter kernel = 3 x 3	0.8399	Blurring kernel = 3 x 3	0.8531
	Median filter kernel = 7 x 7	0.8518	Gaussian filter kernel = 7 x 7	0.8309	Blurring kernel = 7 x 7	0.8520

Tables 6-8 present the results of surgical instrument segmentation under various image processing distortions using three different datasets: JIGSAWS, Endovis, and Endoscape 2023. The segmentation's performance is evaluated using the Intersection over Union (IoU) metric, which measures the overlap between the predicted and ground-truth segmentation masks. Table 6 shows the segmentation performance on the JIGSAWS dataset under different types of image processing attacks where adding different levels of salt and pepper noise slightly decreases the IoU as the noise level increases, with the highest IoU (0.8570) at 0.001 and the lowest (0.7963) at 0.05. On the other hand, the IoU decreases when the image is distorted with JPEG compression. For instance, the IoU drops from 0.8528 at a quality factor of 90 to 0.7524 at a quality factor of 50, indicating that higher compression significantly degrades segmentation performance. Also, the application of different filters (median, Gaussian, and blurring) with different kernel sizes (3x3 and 7x7) shows that the segmentation is quite robust to these distortions, maintaining high IoU values, particularly with a 3x3 kernel size where the IoU remains above 0.9580.

Table 7 presents the results for the Endovis dataset. Like the JIGSAWS dataset, segmentation performance slightly decreases with increased noise levels. The performance against JPEG compression shows a noticeable drop in IoU with higher compression (lower quality factors), from 0.7898 with a quality factor of 90 down to 0.7504 at 50. The results against image filtering indicate that the IoU values are lower than the JIGSAWS dataset.

Finally, Table 8 shows the performance of the Endoscope 2023 dataset; the IoU slightly improves against salt and pepper noise adding, reaching the highest IoU of 0.8657. However, at the highest noise level (0.05), the IoU slightly decreases to 0.8653. Like the other datasets, the IoU decreases as the JPEG compression quality factor is reduced.

The IoU values are generally robust across different filters, with slight variations. The median filter (3x3 kernel) has the best performance at an IoU of 0.8658, while the Gaussian filter (7x7 kernel) has the lowest at an IoU of 0.8309.

Across all datasets, segmentation performance degrades as the salt and pepper noise level increases or as JPEG compression quality decreases, which is expected due to the increased image distortion. The segmentation algorithms are robust to filtering distortions, particularly with smaller kernel sizes. This suggests the model can handle slight smoothing or blurring without significantly compromising accuracy. The JIGSAWS dataset generally shows higher IoU values under distortion than Endovis and Endoscape, indicating that the segmentation model might be better suited or trained for this specific dataset.

These tables and their corresponding analysis highlight the robustness and limitations of the surgical instrument segmentation model when subjected to various common image processing distortions. Understanding these effects is critical for improving the reliability of segmentation algorithms in practical, real-world scenarios where images may undergo different types of preprocessing.

Algorithm comparison

In this section, the proposed surgical instruments segmentation and detection approach is compared with 7 different methods presented in the literature. Which a resume of this comparison is showcase in Table 5.

Table 9 showcases the efficacy of the proposed method for surgical instrument segmentation, offering an efficient and effective solution. Unlike methods reliant on neural networks, the K-means algorithm can discern image patterns and structures autonomously, eliminating the need for pre-labeled data and making it ideal for medical environments requiring surgical instrument detection. Furthermore, the K-means algorithm demonstrates its adaptability by effectively handling changes in camera positioning and environmental illumination, thereby reducing segmentation errors. This robustness sets it apart from other methods that may require more complex supervised information or parameter adjustments, making it a practical and reliable system. The effectiveness of our proposed method is underscored by its comparison with existing research, affirming that the unsupervised approach based on K-means not only simplifies segmentation but also delivers superior efficiency and accuracy in identifying surgical instruments within medical images.

The Table 9 compares different surgical instrument segmentation methods assessed using metrics such as IoU, precision, recall, and F1 Score. Although comparisons have been made with existing methods, it is crucial to highlight that the proposed method outperforms other approaches, especially regarding clinical applicability and computational efficiency.

Unlike complex methods such as supervised deep neural networks, the proposed method uses an auto-supervised segmentation techniques such as K-means and Otsu image binarization for image ground truth generation and a neural network for the generalization of the image segmentation. It is significantly less intensive regarding the requirements for human time to create each image's ground truth. This stage demonstrates a high accuracy for image ground truth creation. Subsequently, it generates efficient neural network training to generalize characteristics and reduce processing time to segment surgical instruments in different scenarios, as in the case of the three datasets. This characteristic makes it a viable option for applications or in resource-constrained environments.

TABLE 9. Literature algorithms comparison with the proposed segmentation and detection approach.

Method	Technique	Metric	Dataset
Allan, <i>et al.</i> ^[20]	Optical flow	Precision = 0.874 Recall = 0.93 F1 Score = 0.898	Ex vivo study from da Vinci Research Kit robotic system ^[10] .
Yu, <i>et al.</i> ^[22]	Neural network: Hollistic Unet	Accuracy = 0.9156 IoU = 0.8645 Dice = 0.9220	Robotic Instrument Segmentation Challenge ^[28] .
Jha, <i>et al.</i> ^[19]	Neural network: DDANet	Recall = 0.8703 Precision = 0.9348 F2 score = 0.8613 Accuracy = 0.9897 Dice = 0.8739 IoU = 0.8739	Robotic Instrument Segmentation Challenge ^[28] .
Xue, <i>et al.</i> ^[23]	Atrous Spatial Pyramid Pooling layer and Convolutional Block Attention Module CBAM	IoU = 0.861 Accuracy = 0.885	No public dataset, the dataset consists in 7 common surgical instruments with a data augmentation.
Wang, <i>et al.</i> ^[21]	Random forest	IoU = 0.9481	Not mention.
Baby, <i>et al.</i> ^[24]	MaskRCNN	IoU = 72.54	Robotic Instrument Segmentation Challenge ^[28] .
Streckert, <i>et al.</i> ^[25]	GAN, SegNet	Dataset Endovio: IoU = 91.21 Synthetic dataset: IoU = 89.55	Robotic Instrument Segmentation Challenge ^[28] .
Proposed Method	Unsupervised segmentation: Machine learning K-means and Otsu image binarization for ground truth creation and FCN model for the generalization of the instrument segmentation	IoU = 0.9641 Acc = 0.9689 Precision = 0.9372 Recall = 0.358 F1 Score = 0.9338 MSE = .0310	Trained with and tested with JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS), in addition it is tested with Endovis and Endoscape datasets ^[29]

In clinical applicability, the proposed method offers a high accuracy of 0.9372 and an IoU of 0.9641, placing it above several methods using advanced neural networks or optical flow techniques. For example, Allan *et al.* report an F1 Score of 0.898 using optical flow in an ex vivo study, while our method achieves an F1 Score of 0.9338 in the primary dataset on the JIGSAWS dataset, which includes surgical tasks in more diverse and challenging environments.

On the other hand, while methods such as the one proposed by Yu *et al.*, with a holistic Unet network, present good precision (0.9220) and an acceptable IoU (0.8645), the proposed method achieves a better balance between precision and recall. Although the proposed method's recall could be higher, this approach could be improved with additional adjustments or combinations with other techniques, but it already demonstrates superiority in terms of overall accuracy. Another aspect to highlight is the proposed method's simplicity compared to approaches such as GAN or SegNet, which require extensive training and the use of synthetic data sets. The proposed method uses a

more direct approach and is less prone to overfitting, maintaining robustness critical in segmenting surgical instruments in various conditions.

Despite being an auto-supervised technique, the proposed approach has a higher generalization ability on different datasets without the need for retraining. The proposed method improves performance metrics and offers clear advantages in practical applicability and computational efficiency. This makes it ideal for implementation in surgical support systems, where speed, precision, and simplicity are essential for clinical adoption.

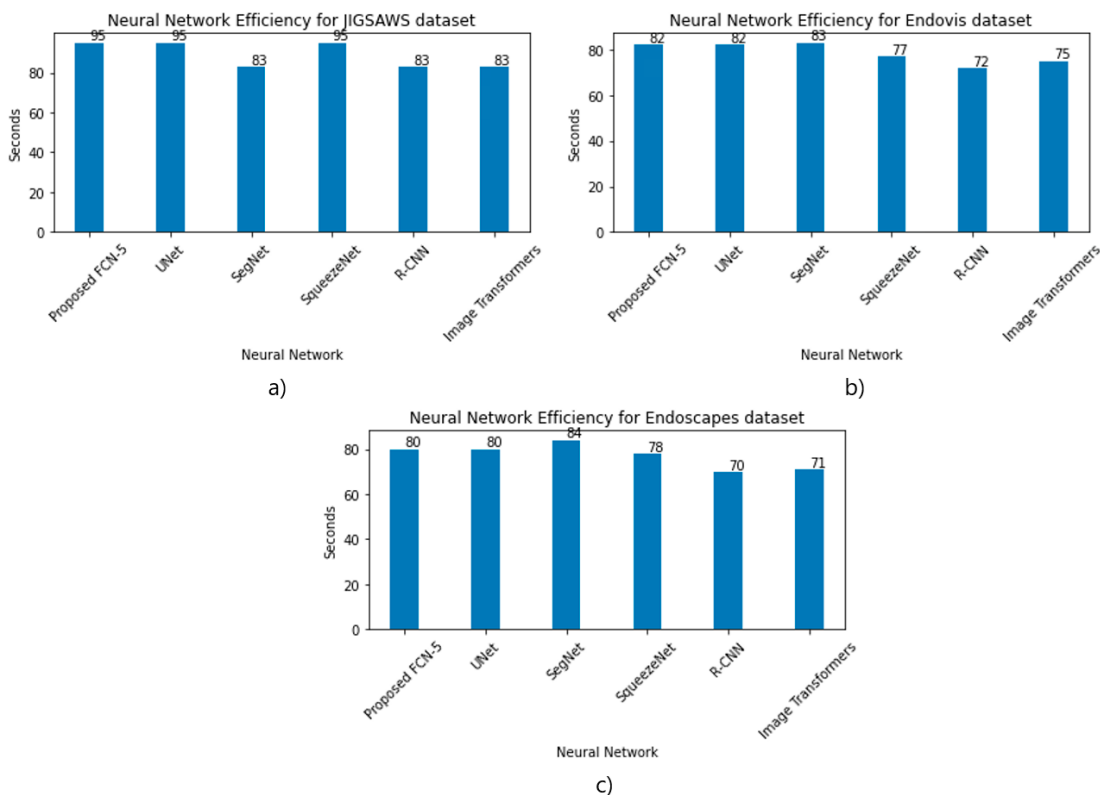


FIGURE 10. Neural networks efficiency for different dataset using JIGSAWS surgical instruments training.

Figure 10 illustrates the efficiency of surgical instrument segmentation using the JIGSAWS database to train different neural networks. Among the models evaluated, UNet and FCN-5 perform better in segmenting these instruments, especially in the Knot-Tying and Needle-Passing tasks. The IoU index measures high efficiency. This algorithm ensures that the surgical instrument's segmentation in the images leads to accurate segmentation without significant errors. This finding underscores the K-means algorithm's effectiveness in image labeling, which contributes to the development of precise segmentation systems. Moreover, the practical application of pre-training with the JIGSAWS database generalizes the surgical instrument segmentation with the neural network training for other tasks. This approach was applied to the EndoVis database; it achieved an 82 % efficiency in recognizing specific surgical instruments used in endoscopies. This adaptability generates an accurate segmentation in diverse contexts, showcasing the robustness of the method.

Combining the neural network with automatic label generation and pre-training with JIGSAWS has proven to be an effective strategy for accurately and efficiently segmenting surgical instruments, both in specific tasks

and broader applications like endoscopies. In addition, the results in the Figure 10 demonstrate that the proposed neural network outperforms other existing models in efficiency, including those based on Image Transformers. This superiority is observed in the segmentation capacity of surgical instruments, especially under controlled conditions. However, to maximize their applicability, the Image Transformers must be adjusted and manipulated to improve their efficiency in complex surgical scenarios. Future research should focus on optimizing these models, seeking a balance between accuracy and efficiency so that they can adapt to a wider variety of imaging conditions and clinical scenarios.

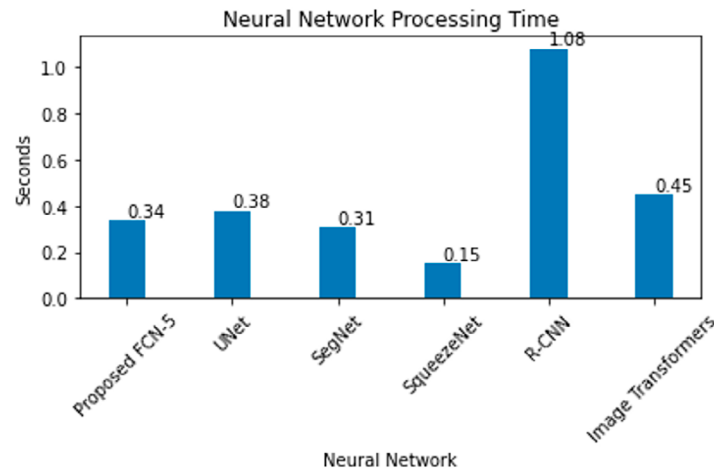


FIGURE 11. Neural networks efficiency for different dataset using JIGSAWS surgical instruments training.

Figure 11 shows the processing time of five different neural networks used for the segmentation of surgical instruments, where SqueezeNet has the fastest processing time with 0.15 seconds, indicating that it is the most efficient network in terms of speed, while SegNet (0.31 seconds), FCN-5 (0.34 seconds), and UNet (0.38 seconds) have moderate processing times, with SegNet being slightly faster than FCN-5 and UNet; On the other hand, RCNN has the longest processing time with 1.08 seconds, which could be a disadvantage in applications where speed is crucial, although SqueezeNet is the fastest network, since UNet and FCN-5 have a higher efficiency in terms of precision in the segmentation of surgical instruments, demonstrating that, despite not being the fastest, they offer a more precise and reliable segmentation, essential in surgical applications. Figure 11 highlights the importance of considering computational cost and accuracy when selecting a neural network for surgical instrument segmentation.

To reduce the computational time without decreasing the accuracy of the proposed system, an adaptive image resolution processing could be implemented, which begins with lower resolution analysis, increases detail only where needed, and can balance speed and accuracy, such as the Unit model. This technique can be used on the ground truth generation using some image descriptors for image region of interest detection where the surgical instrument is located. Multi-scale analysis is another approach that processes images at multiple resolutions simultaneously, allowing for quicker feature identification. Additionally, developing a hybrid model, where a lightweight model performs initial segmentation and is combined with a neural network, such as a GAN model, to refine critical regions, reducing the processing time while maintaining high accuracy.

Discussion

Our research focuses on developing a method for image segmentation of surgical instruments using the unsupervised K-means algorithm to automatically generate the ground truth for training neural networks. The results demonstrate high precision and accuracy in the detection and segmentation of surgical instruments, with a minimum error of 0.0297 and a precision of 0.9702 when evaluating the labels generated for training the neural networks. This approach accurately identifies patterns for detecting surgical instruments despite image shape, size, and texture changes.

The adaptability of the K-means algorithm to changes in lighting and camera position is a key advantage, allowing this algorithm to be used in different surgical environments. The results automate the detection and segmentation of instruments for the generation of ground truth. The proposed method demonstrates superior accuracy, efficiency, and adaptability compared with other existing approaches. Unlike other methods, the presented algorithm in this paper reduces human-machine work to generate training ground truth for systems based on neural networks; the proposed self-supervised approach eliminates the need to create the ground truth of surgical instruments manually. This unique feature makes it ideal for medical environments requiring the detection of surgical instruments.

Table 6 highlights the effectiveness of the proposed method for the segmentation of surgical instruments, offering an efficient solution. Unlike the techniques presented in the literature, the proposed method has the advantage of automatically generating the ground truth for the training of neural networks, which allows generalizing the learning of characteristics of surgical instruments to use them in databases other than those trained to segment surgical instruments precisely. This robustness distinguishes it from other methods that may require more complex monitoring or parameter adjustments, making it a practical and reliable system.

The number of clusters used in the k-means algorithm to detect regions of interest impacts the quality of the segmentation and the processing time. As the clusters increase, more detailed segmentation is allowed as more regions of interest are generated. However, this also increases processing time exponentially due to the computational complexity of the number of clusters. This balance between segmentation quality and processing time is essential in optimizing the segmentation process.

A number of clusters between $N = 5$ and $N = 10$ are ideal, providing high performance (IoU and accuracy) without significantly increasing processing time. The choice of the number of clusters must consider a compromise between the quality of the segmentation and the processing time. The segmentation quality affects the neural network's training, which increases the error when applying the system in a different environment. Some aspects to consider would be the type of system where it is implemented since if it has few processing resources, the processing time would increase even more. Additionally, if the system is implemented in applications different from the evaluation or training of surgeons, environmental changes must be considered when training the neural network. The system is designed to evaluate the surgeon's ability to perform surgery since, generally, even if they are experts, surgeons adjust the environment to the patient's characteristics to increase the surgical procedure's success. Finally, while more clusters can improve segmentation accuracy, especially in images with significant brightness correction, it also significantly increases processing time.

Combining the neural network with the automatic generation of labels generated by the K-means algorithm and pre-training with JIGSAWS has proven to be an effective strategy for accurately and efficiently segmenting surgical instruments in specific tasks and broader applications, such as endoscopies. However, our method has some limitations that will be addressed in future research to improve the segmentation of surgical instruments when using the pretrained neural network. On the other hand, methods should be explored to enhance the precision of segmenting the parts related to the surgical instrument. Implementing our algorithm in current robotic surgery systems could benefit patients and surgeons, automating the detection of surgical instruments and reducing human errors, for example, by detecting and correcting possible collisions between surgical instruments by surgeons.

Based on the results presented in the previous tables, a more developed section on the limitations of the proposed method could include the following points: the proposed method demonstrates strong performance in surgical instrument segmentation under various image distortions, certain limitations should be noted regarding its applicability in different surgical scenarios and imaging conditions. The results from the JIGSAWS, Endovis, and Endoscape datasets indicate that the segmentation performance declines as the salt and pepper noise level increases. The method's performance deteriorates with lower JPEG compression quality factors, with IoU values falling as compression increases (quality factor = 50). In clinical environments where images might be compressed for storage or transmission, this could reduce segmentation accuracy. The method exhibits robustness to filtering techniques such as median, Gaussian, and blurring, the results show variability across different kernel sizes and filter types. For instance, the Endovis dataset displayed lower IoU values under filtering distortions than the JIGSAWS dataset. This inconsistency suggests that the method's robustness might vary depending on the specific characteristics of the dataset or the surgical context, such as the type of instruments or the nature of the surgery. The method performs differently across the JIGSAWS, Endovis, and Endoscape datasets, with the JIGSAWS dataset generally yielding higher IoU values. This disparity indicates that the process may be more finely tuned or optimized for specific datasets. Therefore, its generalizability may be limited across different surgical scenarios, instruments, or imaging devices. This raises concerns about the method's applicability in diverse clinical settings where different datasets with varying characteristics are encountered.

CONCLUSIONS

This paper presents a method for surgical instrument segmentation using the unsupervised K-means algorithm to generate the ground truth necessary to train neural networks automatically. The results demonstrate high precision and accuracy, validating the effectiveness of the auto supervised approach in identifying patterns despite variations in the shape, size, and texture of the images. Combining the K-means algorithm with pretrained neural networks in JIGSAWS has proven an effective strategy, allowing accurate and efficient segmentation in various applications, including endoscopies. Unlike other methods, the algorithm presented in this paper reduces the human effort required to generate training ground truth for neural network-based systems. The proposed self-supervised approach eliminates the need for manual creation of ground truth for surgical instruments. This unique feature makes it ideal for medical environments requiring the detection of surgical instruments. Although some limitations will be addressed in future research, implementing our method in robotic surgery systems promises to improve surgical efficiency and reduce human errors by automating the detection of surgical instruments.

The proposed method shows promise in surgical instrument segmentation, especially under controlled conditions. However, its sensitivity to certain image distortions, variability across datasets, and potential challenges in complex scenarios highlight areas where further refinement and validation are urgently needed. Addressing these limitations will be crucial for enhancing the method's applicability and reliability across a wider range of surgical scenarios and imaging conditions. Training the neural network used in segmentation with images from diverse scenarios and environments, including distorted images, is essential to improve the system image generalization and surgical instrument segmentation. This approach will enhance the model's ability to generalize segmentation tasks and improve efficiency under varying conditions. We can create a more robust and versatile segmentation system by developing an adaptive algorithm capable of adjusting to different environmental factors. Furthermore, continuing the investigation with a more robust neural model is planned as a future research. Combined with the current system, this advanced model will overcome existing limitations and enhance the method's performance, ensuring accurate and reliable surgical instrument segmentation across a broader spectrum of clinical applications.

ACKNOWLEDGEMENT

The authors would like to thank DGAPA for the support provided to carry out this work, through both the UNAM-DGAPA-PAPIIT IT102321 project: "Development of task planning and coordination algorithms for hybrid robots, redundant robots and mobile robots, which interact with each other within a context of an intelligent environment and a cyber-physical system", and the UNAM-DGAPA-PAPIME PE110923 project: "Development of a remote robotics laboratory to implement programming practices of planning and navigation algorithms in physical test benches". D.H-M acknowledges the financial support from CONAHCyT -Mexico through the doctoral scholarship granted ID 857876. R.E.A-A acknowledges the financial support from CONAHCyT -Mexico through the postdoctoral scholarship granted ID 744415. All authors acknowledge the IPN, UNAM and CONAHCyT for the support provided in this research.

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The algorithm, the ground truth, and the resulting segmentation images of this paper are in: <https://github.com/mrg-mex/UMInSe-Unsupervised-surgical-instrument-segmentation>

AUTHOR CONTRIBUTIONS

R. E. A. A. conceptualization, methodology, validation, investigation, and writing original draft; D. H. M. conceptualization, methodology, software, validation, investigation, and writing original draft; M. C. H.

methodology, formal analysis, data curation, writing review and editing, visualization, supervision, project administration and funding acquisition; V. J. G. V. methodology, formal analysis, data curation, writing review and editing, visualization, supervision, project administration and funding acquisition.

REFERENCES

- [1] X. Wang, L. Wang, X. Zhong, C. Bai, X. Huang, R. Zhao, and M. Xia, "Pal-Net: A modified U-Net of reducing semantic gap for surgical instrument segmentation," *IET Image Process.*, vol. 15, no. 12, pp. 2959-2969, 2021, doi: <https://doi.org/10.1049/ipr2.12283>
- [2] S. Nema and L. Vachhani, "Unpaired deep adversarial learning for multi-class segmentation of instruments in robot-assisted surgical videos," *Int. J. Med. Robot.*, vol. 19, no. 4, 2023, no. art. e2514, doi: <https://doi.org/10.1002/rcs.2514>
- [3] E.-J. Lee, W. Plishker, X. Liu, S. S. Bhattacharyya, and R. Shekhar, "Weakly supervised segmentation for real-time surgical tool tracking," *Healthc. Technol. Lett.*, vol. 6, no. 6, pp. 231-236, 2019, doi: <https://doi.org/10.1049/htl.2019.0083>
- [4] N. Ahmidi, L. Tao, S. Sefati, Y. Gao, et al., "A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2025-2041, 2017, doi: <https://doi.org/10.1109/TBME.2016.2647680>
- [5] F. Chadebecq, F. Vasconcelos, E. Mazomenos and D. Stoyanov, "Computer Vision in the Surgical Operating Room," *Visc. Med.*, vol. 36, no. 6, pp. 456-462, 2020, doi: <https://doi.org/10.1159/000511934>
- [6] D. Psychogyios, E. Mazomenos, F. Vasconcelos, and D. Stoyanov, "MSDESIS: Multitask Stereo Disparity Estimation and Surgical Instrument Segmentation," *IEEE Trans. Med. Imaging*, vol. 41, no. 11, pp. 3218-3230, 2022 doi: <https://doi.org/10.1109/tmi.2022.3181229>
- [7] C. González, L. Bravo-Sánchez, and P. Arbeláez, "Surgical instrument grounding for robot-assisted interventions," *Comput. Methods Biomech. Biomed. Eng.: Imaging Vis.*, vol. 10, no. 3, pp. 299-307, 2022, doi: <https://doi.org/10.1080/21681163.2021.2002725>
- [8] W. Burton, C. Myers, M. Rutherford, and P. Rullkoetter, "Evaluation of single-stage vision models for pose estimation of surgical instruments," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 18, no. 12, pp. 2125-2142, 2023, doi: <https://doi.org/10.1007/s11548-023-02890-6>
- [9] Y. Jin, Y. Yu, C. Chen, Z. Zhao, P.-A. Heng, and D. Stoyanov, "Exploring Intra- and Inter-Video Relation for Surgical Semantic Scene Segmentation," *IEEE Tran. Med. Imaging*, vol. 41, no. 11, pp. 2991-3002, 2022, doi: <https://doi.org/10.1109/TMI.2022.3177077>
- [10] M. Attia, M. Hossny, S. Nahavandi, and H. Asadi, "Surgical tool segmentation using a hybrid deep CNN-RNN auto encoder-decoder," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 2017, pp. 3373-3378, doi: <https://doi.org/10.1109/SMC.2017.8123151>
- [11] D. Papp, R. N. Elek, and T. Haidegger, "Surgical Tool Segmentation on the JIGSAWS Dataset for Autonomous Image-based Skill Assessment," 2022 IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC), Reykjavík, Iceland, 2022, doi: <https://doi.org/10.1109/ICCC202255925.2022.9922713>
- [12] M. Daneshgar Rahbar and S. Z. Mousavi Mojab, "Enhanced U-Net with GridMask (EUGNet): A Novel Approach for Robotic Surgical Tool Segmentation," *J. Imaging*, vol. 9, no. 12, 2023, no art. 282, doi: <https://doi.org/10.3390/jimaging9120282>
- [13] E. Colleoni, D. Psychogyios, B. Van Amsterdam, F. Vasconcelos, and D. Stoyanov, "SSIS-Seg: Simulation-Supervised Image Synthesis for Surgical Instrument Segmentation," *IEEE Trans. Med. Imaging*, vol. 41, no. 11, pp. 3074-3086, 2022, doi: <https://doi.org/10.1109/tmi.2022.3178549>
- [14] P. Deepika, K. Udupa, M. Beniwal, A. M. Uppar, V. Vikas, and M. Rao, "Automated Microsurgical Tool Segmentation and Characterization in Intra-Operative Neurosurgical Videos," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, Glasgow, Scotland, United Kingdom, 2022, pp. 2110-2114, doi: <https://doi.org/10.1109/EMBC48229.2022.9871838>
- [15] G. Leifman, A. Aides, T. Golany, D. Freedman, and E. Rivlin, "Pixel-accurate Segmentation of Surgical Tools based on Bounding Box Annotations," 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 2022, pp. 5096-5103, doi: <https://doi.org/10.1109/ICPR56361.2022.9956530>
- [16] R. Mishra, A. Thangamani, K. Palle, P. V. Prasad, B. Mallala, and T.R. V. Lakshmi, "Adversarial Transfer Learning for Surgical Instrument Segmentation in Endoscopic Images," 2023 IEEE International Conference on Paradigm Shift in Information Technologies with Innovative Applications in Global Scenario (ICPSITIAGS), Indore, India, 2023, pp. 28-34, doi: <https://doi.org/10.1109/ICPSITIAGS59213.2023.10527520>
- [17] A. Lou, K. Tawfik, X. Yao, Z. Liu, and J. Noble, "Min-Max Similarity: A Contrastive Semi-Supervised Deep Learning Network for Surgical Tools Segmentation," *IEEE Trans. Med. Imaging*, vol 42, no. 10, pp. 2023, doi: <https://doi.org/10.1109/TMI.2023.3266137>

- [18] E. Colleoni and D. Stoyanov, "Robotic Instrument Segmentation With Image-to-Image Translation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 935-942, 2021, doi: <https://doi.org/10.1109/LRA.2021.3056354>
- [19] D. Jha, S. Ali, N. K. Tomar, M. A. Riegler, and D. Johansen, "Exploring Deep Learning Methods for Real-Time Surgical Instrument Segmentation in Laparoscopy," 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athens, Greece, 2021, doi: <https://doi.org/10.1109/BHI50953.2021.9508610>
- [20] M. Allan, S. Ourselin, D. J. Hawkes, J. D. Kelly and D. Stoyanov, "3-D Pose Estimation of Articulated Instruments in Robotic Minimally Invasive Surgery," *IEEE Trans. Med. Imaging*, vol. 37, no. 5, pp. 1204-1213, pp. 1204-1213, 2018, doi: <https://doi.org/10.1109/tmi.2018.2794439>
- [21] L. Wang, C. Zhou, Y. Cao, R. Zhao, and K. Xu, "Vision-Based Markerless Tracking for Continuum Surgical Instruments in Robot-Assisted Minimally Invasive Surgery," *IEEE Robot. Autom. Lett.*, vol. 8, no. 11, pp. 7202-7209, 2023, doi: <https://doi.org/10.1109/LRA.2023.3315229>
- [22] L. Yu, P. Wang, X. Yu, Y. Yan, and Y. Xia, "A Holistically-Nested U-Net: Surgical Instrument Segmentation Based," *J. Digit. Imaging*, vol. 33, no. 2, pp. 341-347, 2020, doi: <https://doi.org/10.1007%2Fs10278-019-00277-1>
- [23] M. Xue and L. Gu, "Surgical instrument segmentation method based on improved MobileNetV2 network," 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT), Changsha, China, 2021, pp. 744-747, doi: <https://doi.org/10.1109/ISCIPT53667.2021.00157>
- [24] B. Baby, D. Thapar, M. Chasmai, T. Banerjee, et al., "From Forks to Forceps: A New Framework for Instance Segmentation of Surgical Instruments," 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, doi: <https://doi.org/10.1109/WACV56688.2023.00613>
- [25] T. Streckert, D. Fromme, M. Kaupenjohann and J. Thiem, "Using Synthetic Data to Increase the Generalization of a CNN for Surgical Instrument Segmentation," 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 2023, pp. 336-340, doi: <https://doi.org/10.1109/IDAACS58523.2023.10348781>
- [26] Y. Yamada, J. Colan, A. Davila, and Yasuhisa Hasegawa, "Task Segmentation Based on Transition State Clustering for Surgical Robot Assistance," 2023 8th International Conference on Control and Robotics Engineering (ICCRE), Niigata, Japan, 2023, pp. 260-264, doi: <https://doi.org/10.1109/ICCRE57112.2023.10155581>
- [27] Z. Zhang, B. Rosa, and F. Nageotte, "Surgical Tool Segmentation Using Generative Adversarial Networks With Unpaired Training Data," *IEEE Robot. Autom. Lett.*, vol 6, no. 4, pp. 6266-6273, 2021, doi: <https://doi.org/10.1109/LRA.2021.3092302>
- [28] A. Qayyum, M. Bilal, J. Qadir, M. Caputo, et al., "SegCrop: Segmentation-based Dynamic Cropping of Endoscopic Videos to Address Label Leakage in Surgical Tool Detection," 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 2023, doi: <https://doi.org/10.1109/ISBI53787.2023.10230822>
- [29] Y. Gao, S. Vedula, C. E. Reiley, N. Ahmidi, et al., "JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling," *Modeling and Monitoring of Computer Assisted Interventions (M2CAI)*, Boston, United State, 2014.
- [30] L. Maier-Hein, S. Mersmann, D. Kondermann, S. Bodenstedt, et al., "Can Masses of Non-Experts Train Highly Accurate Image Classifiers? A crowdsourcing approach to instrument segmentation in laparoscopic images," *Med. Image Comput. Comput. Assist. Interv.*, vol. 17, pp. 438-445, 2014, doi: https://doi.org/10.1007/978-3-319-10470-6_55
- [31] A. Murali, D. Alapatt, P. Mascagni, A. Vardazaryan, et al., "The Endoscapes Dataset for Surgical Scene Segmentation, Object Detection, and Critical View of Safety Assessment: Official Splits and Benchmark," 2023, arXiv:2312.12429, doi: <https://doi.org/10.48550/arXiv.2312.12429>
- [32] M. Ju, D. Zhang, and Y. J. Guo, "Gamma-Correction-Based Visibility Restoration for Single Hazy Images," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 1084-1088, 2018, doi: <https://doi.org/10.1109/LSP.2018.2839580>
- [33] R. Silpasai, H. Singh, A. Kumarl and L. Balyan, "Homomorphically Rectified Tile-wise Equalized Adaptive Gamma Correction for Histopathological Color Image Enhancement," 2018 Conference on Information and Communication Technology (CICT), Jabalpur, India, 2018, doi: <https://doi.org/10.1109/INFOCOMTECH.2018.8722364>
- [34] H. Ye, S. Yan and P. Huang, "2D Otsu image segmentation based on cellular genetic algorithm," 9th International Conference on Communication Software and Networks (ICCSN), Guangzhou, China, 2017, pp. 1313-1316, doi: <https://doi.org/10.1109/ICCSN.2017.8230322>
- [35] P. Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, 1st Ed. Edinburgh, United Kingdom: Cambridge University Press, 2012.
- [36] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python*, United States of America: O`Reilly, 2017.
- [37] P. Yin, R. Yuan, Y. Cheng and Q. Wu, "Deep Guidance Network for Biomedical," *IEEE Access*, vol. 8, pp. 116106-116116, 2020, doi: <https://doi.org/10.1109/ACCESS.2020.3002835>

- [38] I. El rube', "Image Color Reduction Using Progressive Histogram Quantization and K-means Clustering," 2019 International Conference on Mechatronics, Remote Sensing, Information Systems and Industrial Information Technologies (ICMRSISIT), Ghana, 2020, doi: <https://doi.org/10.1109/ICMRSISIT46373.2020.9405957>
- [39] R. E. Arevalo-Ancona and M. Cedillo-Hernandez, "Zero-Watermarking for Medical Images Based on Regions of Interest Detection using K-Means Clustering and Discrete Fourier Transform," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 6, 2023, doi: <https://dx.doi.org/10.14569/IJACSA.2023.0140662>
- [40] P. Sharma, "Advanced Image Segmentation Technique using Improved K Means Clustering Algorithm with Pixel Potential," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Wanknaghat, India, 2020, pp. 561-565, doi: <https://doi.org/10.1109/PDGC50313.2020.9315743>