

Aplicación y Comparativa de Modelos Estado del Arte en Segmentación Semántica para Imágenes Laparoscópicas

Application and Comparison of State-of-the-Art Models in Semantic Segmentation for Laparoscopic Images

Diana J. Torres-López¹, Daniel A. Gutiérrez-Jiménez², Sebastián Salazar-Colores³

¹ Universidad de Guanajuato.

^{2,3} Centro de Investigaciones en Óptica (CIO), León 37150, México.

dj.torreslopez@ugto.mx¹, daniel.gutierrez@cio.mx², sebastian.salazar@cio.mx³.

Resumen

En el ámbito médico actual, la laparoscopia emerge como una técnica quirúrgica de mínima invasión, esencial tanto en el diagnóstico como en el tratamiento de patologías intraabdominales. A pesar de su revolucionario impacto, al reducir traumatismos y optimizar la visualización mediante el uso del laparoscopio, esta modalidad enfrenta desafíos inherentes, particularmente en su integración con sistemas robóticos avanzados, como el sistema Da Vinci. Uno de los retos más críticos es la precisa identificación y seguimiento de las herramientas quirúrgicas, crucial para garantizar una operación exitosa y segura. Con el advenimiento de la inteligencia artificial, y en particular de la visión por computadora, se han propuesto soluciones basadas en el aprendizaje profundo para la detección y rastreo de instrumentos en entornos laparoscópicos. No obstante, aún se observan limitaciones, especialmente en la segmentación precisa de instrumentos frente a tejidos adyacentes. Ante este escenario, el presente estudio propone investigar la eficacia de algoritmos de vanguardia en segmentación semántica de imágenes para solventar dichas limitaciones. Más allá de un análisis comparativo de algoritmos actuales, este trabajo aspira a sentar precedentes para el desarrollo de métodos de segmentación especializados para imágenes laparoscópicas. La segmentación precisa es imperativa en la laparoscopia, facilitando la identificación de áreas críticas, optimizando maniobras quirúrgicas y reduciendo riesgos intraoperatorios. A través de esta investigación se busca no solo potenciar la laparoscopia, sino también ofrecer una herramienta integral que aumente la seguridad y eficiencia, y que permita una disminución de los costos asociados a tecnologías médicas de punta.

Palabras clave: Inteligencia artificial, Imágenes laparoscópicas, Segmentación semántica, Aprendizaje profundo, Visión por computadora en medicina.

Introducción

La laparoscopia, también conocida como cirugía laparoscópica o cirugía mínimamente invasiva, es un procedimiento médico fundamental utilizado tanto para el diagnóstico como para el tratamiento de diversas afecciones dentro del cuerpo humano. En contraste con la cirugía tradicional, que involucra incisiones considerables, la laparoscopia se vale de instrumentos especializados y un laparoscopio, una cámara pequeña, para acceder al área de interés mediante pequeñas incisiones en la piel. El laparoscopio, un tubo delgado y flexible equipado con una luz y una cámara en su extremo, se introduce a través de una de las incisiones, transmitiendo imágenes en tiempo real a un monitor en el quirófano. Esto permite al cirujano obtener una visión clara del interior del cuerpo humano. Los demás instrumentos se insertan a través de incisiones adicionales para realizar procedimientos quirúrgicos que abarcan desde la extirpación de órganos enfermos hasta la toma de muestras de tejido, la reparación de tejidos dañados o procedimientos diagnósticos.

La laparoscopia ofrece numerosas ventajas en comparación con la cirugía tradicional, incluyendo una reducción significativa del traumatismo, el dolor posoperatorio y el tiempo de recuperación (*Bartoli et al., 2012*). Este enfoque se ha convertido en una técnica comúnmente empleada en cirugías abdominales, exenteración pélvica, reparación de aneurisma aórtico abdominal, histerectomía y diversas áreas de la medicina, abordando una amplia variedad de afecciones y procedimientos.

Hoy en día, la cirugía robótica asistida por laparoscopia ha ganado una gran popularidad y se ha convertido en un campo en constante crecimiento en la medicina. Los sistemas robóticos como el Da Vinci, Senhance Surgical System y los simuladores como el Mimic dV-Trainer y el RoSS surgical simulator (*Chahal et al., 2022*) se utilizan para una amplia gama de aplicaciones médicas y quirúrgicas y entrenamiento. En las cirugías asistidas por robot es importante la localización de las herramientas quirúrgicas, ya que nos puede dar una mejor interpretación de la interacción entre herramientas y tejido y de esta forma evitar lesiones que pueden llegar a ser graves o fatales.

La inteligencia artificial (IA) es una disciplina que está generando un impacto significativo gracias a su amplia aplicabilidad y flexibilidad. Ha demostrado ser una herramienta poderosa que aborda tanto problemas de índole cotidiana como cuestiones altamente complejas en campos científicos e industriales de alto nivel (*Valbuena-Pardal, 2022*).

Esencialmente, la identificación de objetos representa una tarea central en uno de los campos más estudiados en inteligencia artificial: la visión por computadora. Esta disciplina se dedica a la interpretación y comprensión del contenido visual presente en imágenes y vídeos. Que, junto con el aprendizaje profundo y su capacidad para aprender características y patrones complejos de datos visuales, desempeñan un papel fundamental en aplicaciones que engloban desde la extracción de características específicas en imágenes, reconocimiento de patrones, clasificación, detección y segmentación para la identificación de objetos individuales (*Minaee et al. 2022*).

En el pasado, para abordar la detección sin utilizar segmentación, se empleaban sensores externos o soluciones mecánicas. Los sensores externos, en su mayoría, eran rastreadores ópticos altamente precisos, costosos y voluminosos que podían seguir marcadores específicos en el espacio 3D. Aunque eran altamente efectivos para ubicar estos marcadores y su uso era sencillo, estos métodos tienen dos desventajas fundamentales: su elevado costo y la necesidad de modificar las herramientas quirúrgicas para acomodar los marcadores que debían ser rastreados. Por ejemplo, el sistema Da Vinci utiliza codificadores de motores de alta precisión para localizar la punta de la herramienta durante la operación, lo que requiere el desarrollo de un sistema fabricado con gran precisión, lo que puede resultar muy costoso (*Souipas et al., 2023*).

Actualmente, se cuenta con una amplia disponibilidad de videos que documentan procedimientos quirúrgicos. Estos recursos resultan muy valiosos para realizar análisis minuciosos y evaluar la habilidad de los cirujanos, además de identificar posibles errores técnicos. Su relevancia trasciende a la comunidad médica en general, ya que posibilitan la observación de la manipulación de los instrumentos quirúrgicos.

En este contexto, la detección automatizada de instrumentos quirúrgicos desempeña un papel importante, tanto para el análisis retrospectivo como en la mejora del proceso quirúrgico en sí. Con el propósito de abordar esta cuestión, se ha enfocado el esfuerzo en el desarrollo de redes neuronales profundas, particularmente, en las redes neuronales convolucionales. Estas últimas han demostrado ser particularmente efectivas en la tarea de segmentar objetos, lo que representa un avance de gran relevancia en este campo.

Si bien se ha trabajado en la detección y seguimiento de instrumentos en videos, lo cual resulta útil en tiempo real, se ha enfrentado a limitaciones en cuanto a la precisión de la segmentación individual y cuando se concentra en la segmentación de instrumentos en imágenes individuales (comúnmente empleando técnicas de segmentación semántica), surge el desafío de la alta similitud entre los instrumentos y el fondo, lo cual complica la tarea y motiva la búsqueda de mejoras en este aspecto. (*Kletz, 2019*).

Es importante enfatizar la necesidad de examinar minuciosamente la incorporación de datos, ya que esto conduce a la obtención de resultados más precisos y confiables al utilizar modelos de aprendizaje profundo. La inclusión de imágenes segmentadas, en particular, puede potenciar la capacidad de estos modelos para llevar a cabo tareas específicas (*Leyva-López et al., 2023*).



Dado el éxito de las estrategias de visión asistida por computadora en el campo médico y el constante avance del aprendizaje profundo y sus aplicaciones, como la segmentación, se presenta una oportunidad valiosa en el ámbito de la laparoscopia. Esta tecnología no solo ofrece un enfoque más económico, sino también una retroalimentación precisa para los profesionales de la medicina, sin la necesidad de utilizar costosas y complejas maquinarias.

En este contexto, proponemos una investigación centrada en la evaluación de modelos basados en aprendizaje profundo para la segmentación y localización de herramientas quirúrgicas y tejido en imágenes laparoscópicas. Esta aproximación tiene el potencial de revolucionar la laparoscopia al mejorar la precisión de la identificación de órganos y reducir los costos asociados con equipos médicos avanzados, al tiempo que proporciona una valiosa herramienta para los cirujanos en su toma de decisiones clínicas.

Fundamentos teóricos

Laparoscopia

La cirugía laparoscópica representa una aproximación quirúrgica menos invasiva en comparación con la cirugía tradicional, lo que conlleva a una recuperación postoperatoria más rápida. Aunque la cirugía laparoscópica ha experimentado un rápido crecimiento y aceptación para ciertas condiciones, enfrenta desafíos como la necesidad de habilidades manuales más precisas y una duración quirúrgica prolongada. El futuro de esta técnica dependerá de su capacidad para establecerse como un procedimiento quirúrgico estándar en un plazo razonable, así como de la investigación y el desarrollo de nuevas técnicas y tecnologías mínimamente invasivas respaldadas por estudios comparativos que demuestren su viabilidad y perspectivas a largo plazo (*Cuesta, 2000*).

Entre las ventajas de una cirugía laparoscópica encontramos que esta disminuye la necesidad de analgésicos, efectos secundarios, probabilidad de infecciones de heridas, eventos de trombosis venosa profunda y embolia, complicaciones neurológicas. Además, mejora la apariencia estética pues son incisiones muy pequeñas, tiempo de recuperación y visión quirúrgica (*Menzies, 2016*)

Sin embargo, puede enfrentarse a desventajas como la pérdida de retroalimentación táctil y visión estereoscópica, daños a los vasos sanguíneos y el intestino, aparición ocasional de hernias (menos frecuentes), siendo adicionalmente necesaria una aptitud quirúrgica para este tipo de procedimientos. (*Crist & Gadacz, 1993*).

Inteligencia artificial (IA)

La inteligencia artificial se refiere a la capacidad de un sistema informático para llevar a cabo tareas que normalmente requieren de la inteligencia humana, como el razonamiento, el aprendizaje y la toma de decisiones, a menudo involucrando la manipulación de datos en escalas que superan las capacidades humanas. La IA se sustenta principalmente en el aprendizaje automático (Machine learning, "ML" por sus siglas en inglés) y el aprendizaje profundo (Deep learnig "DL" por sus siglas en inglés), que son utilizados para analizar datos, generar predicciones y proyecciones, categorizar objetos, procesar lenguaje natural, ofrecer recomendaciones y llevar a cabo búsquedas inteligentes de información, entre otros usos.

El aprendizaje automático se nutre de vastas cantidades de datos estructurados y tiene la capacidad de adaptarse y mejorar de manera continua, lo que posibilita un aprendizaje constante y permanente.



Mientras que el aprendizaje profundo se centra en la construcción y entrenamiento de redes neuronales artificiales profundas para resolver tareas y problemas complejos. Se ha destacado por que tiene capacidad para abordar problemas en una amplia gama de campos, como: el procesamiento de imágenes, el procesamiento de lenguaje natural, la visión por computadora, la traducción automática, la conducción autónoma, el reconocimiento de voz, análisis de analizar datos médicos complejos, entre otros. (Jiang et. al., 2022).

Redes Neuronales Convolucionales (CNN)

Las redes CNN son un tipo específico de arquitectura de redes neuronales artificiales y técnica específica dentro del campo de la IA, siendo útiles para realizar tareas de visión por computadora y procesamiento de imágenes. Las CNN utilizan capas de convolución que van aprendiendo automáticamente características importantes en imágenes (bordes, formas y patrones) a medida que profundizan en la red. Esto las hace especialmente adecuadas para tareas de procesamiento, clasificación y segmentación de imágenes, así como para la detección de objetos (Zhang et. al., 2021).

Para la segmentación de imágenes y detección de objetos pueden utilizarse arquitecturas de CNN en particular, como lo son: SegNet, U-Net y FCN.

Red Neuronal Completamente Convolutiva (FCN)

La Red Neuronal Completamente Convolutiva (FCN, por sus siglas en inglés) es una arquitectura diseñada para la segmentación semántica de imágenes. A diferencia de las redes convolucionales tradicionales que se enfocan en la clasificación, las FCN se adaptan para asignar etiquetas a cada píxel, logrando una segmentación precisa. Para manejar imágenes de diferentes tamaños, estas redes utilizan capas convolucionales y luego aplican técnicas de "upsampling" o deconvolución para restituir la resolución original de la imagen. Las conexiones de salto en las FCN permiten integrar información de mapas de características de distintas profundidades, combinando detalles finos y contextos más amplios. Las FCN no sólo se aplican a la segmentación semántica, sino también a otras tareas en visión por computadora como detección de objetos. Su ventaja reside en procesar imágenes completas en una sola ejecución, adaptándose eficientemente a diferentes tamaños de entrada.

Arquitectura U-Net

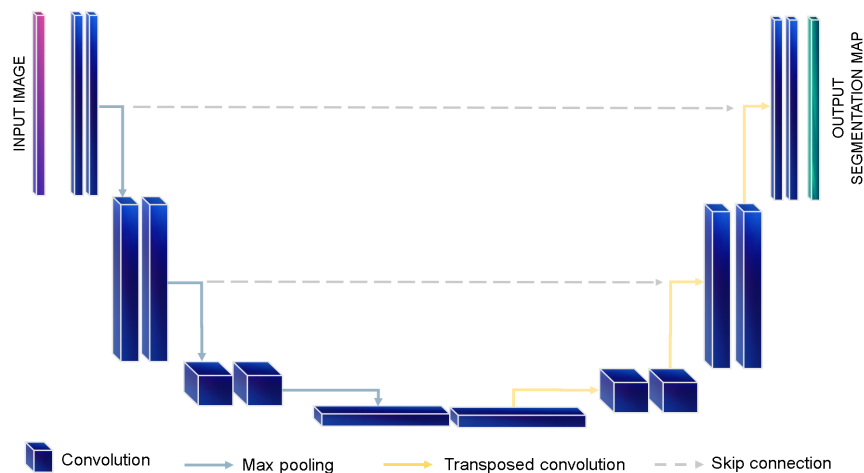


Figura 1. Arquitectura U-Net.
Fuente: Elaboración propia.

La arquitectura U-Net (Figura 1) se ha establecido como un referente en la segmentación semántica de imágenes, particularmente en el ámbito biomédico. Esta estructura se distingue por su forma en "U", compuesta por un camino de contracción (o codificación) seguido de un camino expansivo (o decodificación). En la fase de codificación, las características espaciales de la imagen se reducen progresivamente mediante capas convolucionales y operaciones de "pooling". Sin embargo, la profundidad de los mapas de características aumenta. En la fase de decodificación, la resolución espacial de estos mapas se incrementa mediante operaciones de "upsampling" o deconvolución, mientras que su profundidad disminuye. Lo que realmente distingue a U-Net es la incorporación de conexiones de salto que transfieren información de capas de codificación a capas correspondientes en la decodificación. Estas conexiones aseguran que tanto las características de alta resolución (detalles finos) como las de baja resolución (contexto) se utilicen para producir una segmentación precisa. La eficacia de U-Net en la segmentación precisa de estructuras en imágenes biomédicas ha propiciado su adaptación y uso en múltiples aplicaciones dentro y fuera del ámbito médico.

Sequential Fully Convolutional Networks (SegNet).

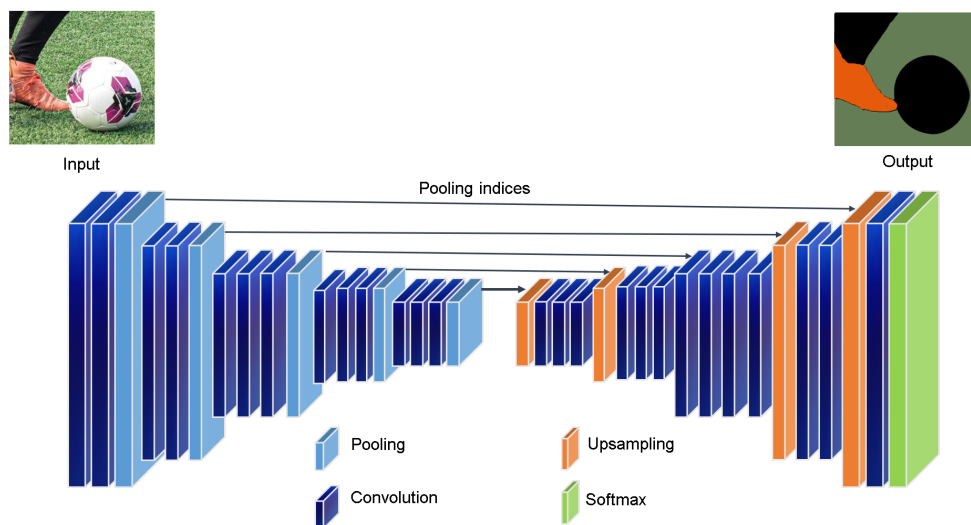


Figura 2. Sequential Fully Convolutional Networks (SegNet).
Fuente: Elaboración propia y Elk Sport (2014).

La arquitectura SegNet (Figura 2) es una propuesta de segmentación semántica basada en redes neuronales profundas, particularmente diseñada para abordar tareas de segmentación en imágenes urbanas. La estructura de SegNet se puede dividir esencialmente en dos partes principales: un codificador y un decodificador. El codificador está compuesto por capas convolucionales similares a las primeras capas de la conocida red VGG16, donde cada capa está seguida por una operación de "pooling" que reduce la dimensionalidad espacial mientras captura las características más prominentes. Durante esta fase, los índices de las posiciones máximas de "pooling" se almacenan para su uso posterior en la fase de decodificación. En el decodificador, estos índices almacenados se utilizan para realizar un "upsampling" de los mapas de características, lo que permite una mejor recuperación de los detalles espaciales de la imagen original. Tras el proceso de decodificación, se obtiene un mapa de segmentación con las mismas dimensiones que la imagen de entrada. La capacidad de SegNet para mantener la resolución espacial mediante el uso de índices de "pooling" la ha posicionado como una opción robusta para segmentación en escenarios donde la precisión de los detalles es crucial.



Métrica de evaluación de imagen: Intersección sobre Unión o IoU

La segmentación, detección y seguimiento de objetos ha sido una de las áreas en investigación más importantes en visión por computadora, para evaluar este tipo de tareas se puede utilizar una de las métricas más importantes en la actualidad, la Intersección sobre Unión (IoU) ó índice Jaccard, la IoU, la cual evalúa que tan cerca está un cuadro delimitado de predicción de la verdad fundamental, es decir, mide la precisión con la que el modelo ha identificado y ubicado el objeto al comparar la región predicha con la región real (Rezatofighi, H. et. al., 2019)

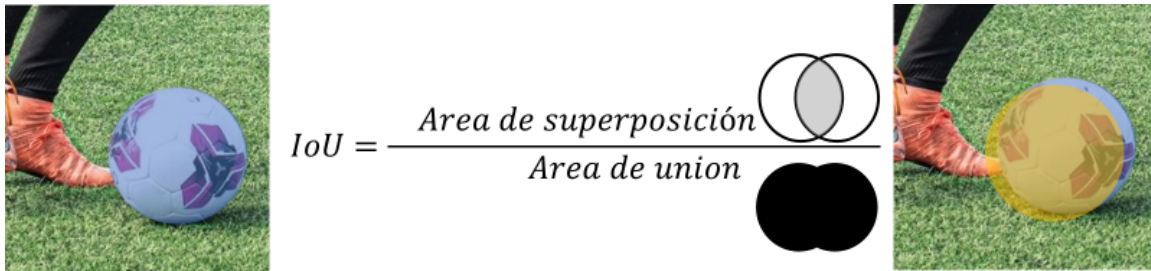


Figura 3. Intersección sobre Unión (IoU).
Fuente: Elaboración propia y Elk Sport (2014).

Se entrena un modelo para generar un área que se ajusta perfectamente alrededor de un objeto. Por ejemplo, en la *Figura 3* tenemos un área azul y un área amarilla. El área azul representa lo real y el área amarilla es la predicción de nuestro modelo. De manera que este modelo busca mejor su predicción, hasta que el área azul y el área amarilla se superpongan perfectamente hasta que el IoU entre los dos cuadros sea igual a 1.

Metodología Propuesta

La investigación se centró en la evaluación de redes neuronales convolucionales (CNN) incorporando un conjunto de datos compuesto por 20,000 imágenes distribuidas en tres categorías distintas. Estas imágenes se sometieron a un proceso de entrenamiento y evaluación en distintos códigos con el objetivo de seleccionar el modelo que tuviera una mejor respuesta en cuanto al nivel de segmentación de instrumentos, tomando en consideración las métricas y resultados de prueba de las imágenes. Para una comprensión más detallada, este proceso se divide en cuatro etapas principales:

- I. **Uso de datos y selección de modelos.** La base de datos consistió en un total de 20,000 imágenes con códigos específicos para segmentación y etiquetado de imágenes, en este caso fueron: "Labels segmentation and translations" y "cholecs80 style" obtenidos de la una colección abierta de conjunto de datos para sistemas quirúrgicos (Pfeiffer et. al., 2019).
- II. **Prueba de redes convolucionales.** Se pusieron a prueba modelos diseñados por Divam Gupta, De manera que el entrenamiento se llevó a cabo en 3 diferentes tipos de arquitectura: unet (vgg_unet, resnet50_unet y mobilenet_unet), fcn (fcn_8 y fcn_32) y segnet (vgg_segnet, resnet50_segnet y mobilenet_segnet).
- III. **Categorización.** Para la evaluación y entrenamiento, el total de imágenes se dividieron en 3 categorías: entrenamiento, validación y prueba, en una proporción de 70%, 10% y 20% respectivamente.
- IV. **Selección.** Al ser evaluadas las bases de datos con los modelos seleccionados, se generaron dos carpetas: la de imágenes originales y la segunda de imágenes con etiquetas (6 etiquetas en total, donde se incluyeron tejidos e instrumental quirúrgico). Posteriormente se entrenaron y evaluaron las redes CNN con las etiquetas, para analizar la precisión para identificar los instrumentos y tejidos en base a la métrica IoU.

Resultados

Resultados cualitativos

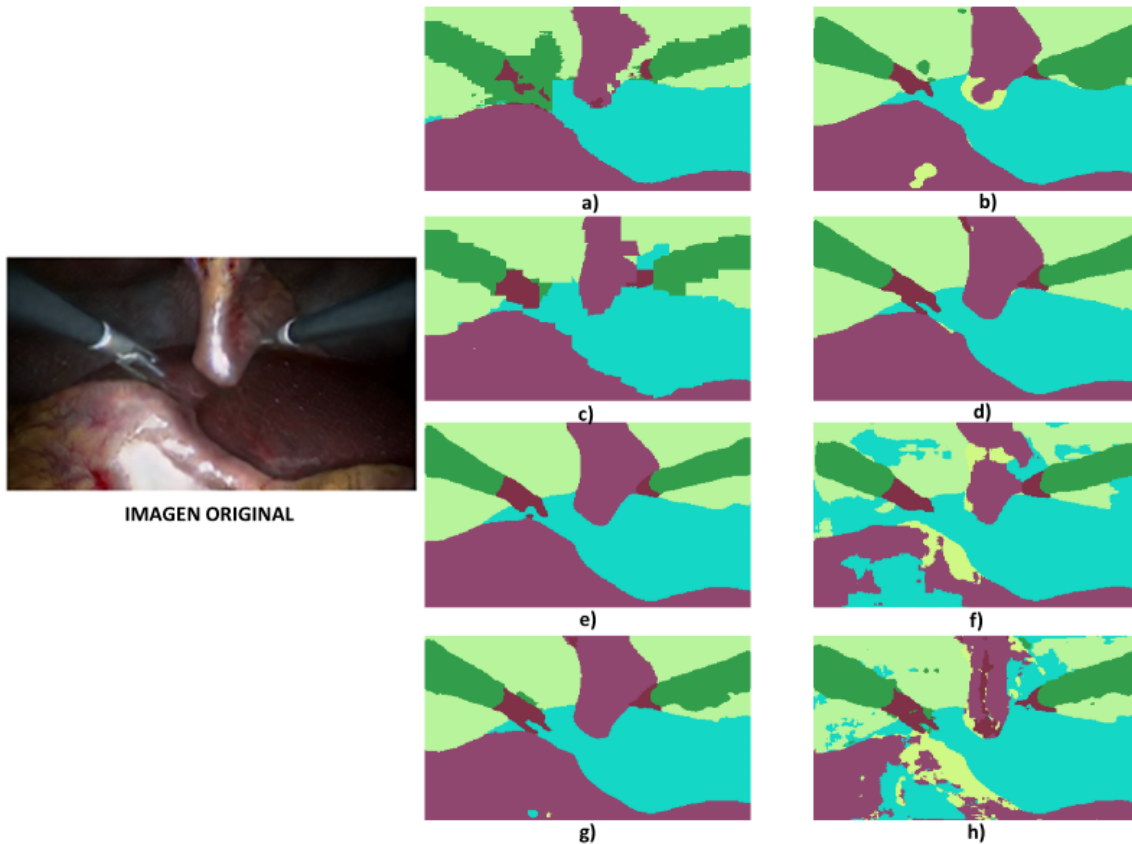


Figura 4. Resultados de la evaluación de cada red neuronal contra la imagen original. a) fcn 8, con métrica promedio de: 0.532. b) resnet 50 segnet, con métrica promedio de: 0.624. c) fcn 32, con métrica promedio de: 0.582. d) resnet 50 unet, con métrica promedio de: 0.731. e) mobilenet segnet, con métrica promedio de: 0.735. f) vgg segnet, con métrica promedio de: 0.557. g) mobilenet unet, con métrica promedio de: 0.710. h) vgg unet, con métrica promedio de: 0.558.

Fuente: Elaboración propia.

Una vez entrenadas las redes neuronales se obtuvieron una serie de imágenes, de las cuales se seleccionó el mejor candidato por cada modelo de red evaluado (ver *Figura 4*), de manera que a primera vista los tres modelos con baja calidad en segmentación e identificación de las etiquetas asignadas corresponden al fcn 8 (*figura 4a*), vgg unet (*figura 4h*) y vgg segnet (*figura 4f*), mientras que los mejores tres resultados son mobile unet (*figura 4g*), mobile segnet (*figura 4e*), y resnet 50 unet (*figura 4d*). Esta interpretación puede justificarse con los resultados cuantitativos.

Resultados cuantitativos

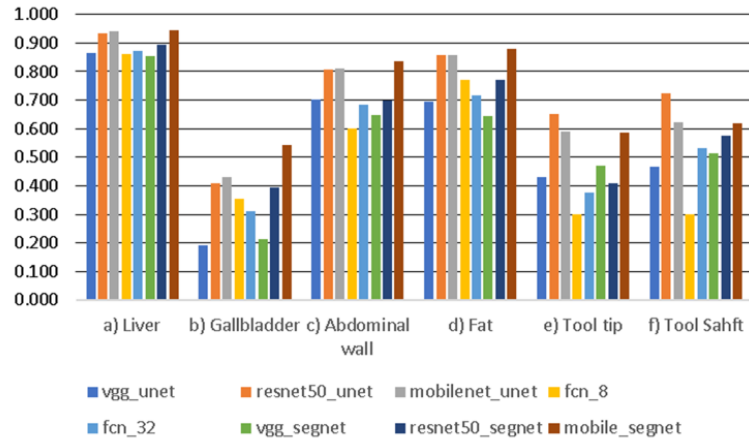


Figura 5. Comparación del desempeño de los métodos propuestos de acuerdo con la IoU de cada código en cada etiqueta asignada. a) "Liver" etiqueta para la zona correspondiente a hígado, b) "Gallbladder" etiqueta para la zona correspondiente a vesícula, c) "abdominal wall" etiqueta para la zona de pared abdominal, d) "fat" etiqueta para la zona de grasa., e) "tool tip" etiqueta para la zona de la punta de la herramienta y f) "tool shaft" etiqueta para la zona del eje de la herramienta.
Fuente: Elaboración propia.

Como se mencionó anteriormente para el procesamiento y entrenamiento de los métodos propuestos, a las imágenes se les asignaron etiquetas previamente, con ayuda de los códigos "Labels segmentation and translations" y "cholecs80 style" y las etiquetas se asignaron de acuerdo a la zona delimitada (en total 6 etiquetas). Posteriormente el entrenamiento se llevó a cabo en 3 diferentes tipos de arquitectura (unet, fcn y segnet para obtener el valor del IoU.

El índice IoU se comparó en cada etiqueta generada por cada código evaluado (Figura 5). Las zonas evaluadas incluyeron áreas como el hígado (Figura 5a), vesícula (Figura 5b), pared abdominal (Figura 5c), grasa (Figura 5d), la punta de la herramienta laparoscópica (Figura 5e) y el eje de la herramienta laparoscópica (Figura 5f). Recordemos que entre más cercano a 1 sea la métrica IoU, esto significará una superposición más acertada al área real delimitada.

De manera que podemos decir que una de las áreas con mejor IoU es el hígado (figura 5a), donde todos los códigos evaluados se mantuvieron en rango mayor a 0.8. Para el caso del área de la vesícula (Figura 5b), el modelo con mejor índice fue al ser evaluada por un código de tipo "mobile_segnet" con un índice mayor a 0.5. En cuanto a la pared abdominal (Figura 5c) y zona grasa (Figura 5d) también tuvieron su mejor nivel de IoU al ser evaluadas por con "mobile_segnet" pero en índices mayores a 0.8.

Finalmente, las herramientas laparoscópicas, punta (Figura 5e) y eje (Figura 5f), los mejores índices obtenidos fueron al ser evaluadas con "resnet50_unet" con índices mayores a 0.6 y 0.7 respectivamente.

Tabla 1. Resultados promedio de las métricas por código y tipo de arquitectura.

Código	\bar{x} por código	\bar{x} por arquitectura
vgg_unet	0.558	0.666
resnet50_unet	0.731	
mobilenet_unet	0.710	
fcn_8	0.532	0.557
fcn_32	0.582	
vgg_segnet	0.557	0.639
resnet50_segnet	0.624	
mobile_segnet	0.735	
		\bar{x} : media aritmética.

Fuente: Elaboración propia.

A fin de tener un panorama más general de los resultados con cada modelo evaluado se compararon los valores promedio o media aritmética por cada código evaluado y tipo de arquitectura (Tabla 1).

De manera que en cuanto a sus índices promedio podemos decir que uno de los resultados más bajos corresponde a la arquitectura de tipo "fcn" y el más alto correspondió a la arquitectura "unet", seguido del modelo "segnet".

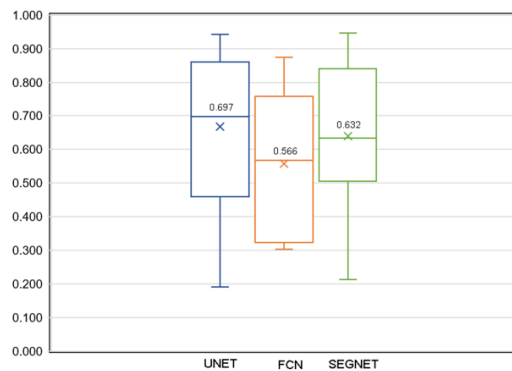


Figura 6. Comparación del desempeño de los métodos propuestos de acuerdo con la IoU obtenida.
 Fuente: Elaboración propia.

Gráficamente pueden verse los desempeños de cada modelo (figura 6), donde se puede confirmar lo anteriormente discutido, en primer lugar, el modelo "unet" alcanzando valores IoU= 0.697 y "segnet" en segundo lugar con IoU= 0.632.

El modelo unet es actualmente uno de los más ampliamente empleados. Recientemente, se ha aplicado exitosamente a la segmentación de imágenes médicas, incluyendo las oculares (Bhattarai et al., 2023), cardíacas y de diversos órganos (Cao et al., 2023). De manera similar, el modelo segnet ha sido evaluado en imágenes de endoscopia laríngea (Laves et al., 2019), con resultados positivos.

Conclusión

A partir de las evaluaciones llevadas a cabo, podemos concluir que la arquitectura U-Net se destacó como la más efectiva para la segmentación semántica de imágenes laparoscópicas. Esta conclusión se basa en una comparación cuantitativa del IoU en diversas áreas de interés, como el hígado, la vesícula, la pared abdominal, la grasa, y las herramientas laparoscópicas. En específico, cabe resaltar que el modelo U-Net sobresalió al exhibir un rendimiento significativamente superior en términos del índice de superposición (IoU) promedio. Asimismo, SegNet presentó resultados competitivos en este contexto. Esta observación refuerza la eficacia demostrada por RCN en la identificación y segmentación de estructuras anatómicas importantes presentes en imágenes laparoscópicas. De manera particular estos hallazgos sugieren que la aplicación de U-Net en la segmentación semántica de imágenes laparoscópicas puede mejorar significativamente la identificación de áreas críticas durante procedimientos quirúrgicos, contribuyendo así al objetivo de mejora en la seguridad y eficiencia de cirugías laparoscópicas.

Referencias

- Bartoli, A., Collins, T., Bourdel, N., Canis, M. (2012). *Computer Assisted Minimally Invasive Surgery: Is Medical Computer Vision the Answer to Improving Laparoscopy?* *Medical Hypotheses*, 79(6), 858–863. <https://doi.org/10.1016/j.mehy.2012.09.007>
- Bhattarai, B., Subedi, R., Gaire, R. R., Vazquez, E., & Stoyanov, D. (2023). *Histogram of Oriented Gradients meet deep learning: A novel multi-task deep network for 2D surgical image semantic segmentation.* *Medical Image Analysis*, 102747. <https://doi.org/10.1016/j.media.2023.102747>
- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2023). *Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation.* *Lecture Notes in Computer Science*, 205–218. <https://arxiv.org/abs/2105.05537>
- Chahal, B., Aydın, A., Ali Amin, M. S., Ong, K., Khan, A., Khan, M. S., Ahmed, K., & Dasgupta, P. (2022). *Transfer of open and laparoscopic skills to robotic surgery: A systematic review.* Published online: November 22, 2022. <https://doi.org/10.1007/s11701-022-01492-9>
- Crist, D. W., & Gadacz, T. R. (1993). *Complications of Laparoscopic Surgery.* *Surgical Clinics of North America*, 73(2), 265–289. [https://doi.org/10.1016/S0039-6109\(16\)45981-5](https://doi.org/10.1016/S0039-6109(16)45981-5)
- Cuesta, M. A. (2000). *Cirugía laparoscópica.* *Cirugía Española*, 68(4), 420–423. <https://www.elsevier.es/es-revista-cirugia-espanola-36-articulo-cirugia-laparoscopica-12567>
- Elk Sport (2014). *Medidas y especificaciones de los balones de fútbol 11.* *Elksport*. <https://elksport.com/blog/medidas-y-especificaciones-de-los-balones-de-futbol-11>
- Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). *Quo vadis artificial intelligence?* *Discover Artificial Intelligence*, 2(1). Springer nature. <https://doi.org/10.1007/s44163-022-00022-8>
- Kletz, S., Schoeffmann, K., Benois-Pineau, J., & Husslein, H. (2019). *Identifying Surgical Instruments in Laparoscopy Using Deep Learning Instance Segmentation.* *IEEE Xplore*. <https://doi.org/10.1109/CBML.2019.8877379>
- Laves, M.-H., Bicker, J., Kahrs, L. A., & Ortmaier, T. (2019). *A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation.* *International Journal of Computer Assisted Radiology and Surgery*, 14(3), 483–492. <https://doi.org/10.1007/s11548-018-01910-0>
- Leyva-López, S., Hernández-Nava, G., Mena-Camilo, E., & Sebastián Salazar-Colores. (2023). *Improving Idiopathic Pulmonary Fibrosis Damage Prediction with Segmented Images in a Deep Learning Model.* *IEEE Conference on Artificial Intelligence (CAI)*, Santa Clara, CA, USA, 2023, pp. 163-165. <https://doi.org/10.1109/CAI54212.2023.00078>
- Menzies, D. (2016). *Medical Malpractice: Laparoscopic Surgery.* *Encyclopedia of Forensic and Legal Medicine: Second Edition*, 432–437. <https://doi.org/10.1016/B978-0-12-800034-2.00263-9>



- Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., & Terzopoulos, D. (2021). *Image Segmentation Using Deep Learning: A Survey*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. <https://doi.org/10.1109/TPAMI.2021.3059968>
- Pfeiffer, M., Funke, I., Robu, M. R., Bodenstedt, S., Strenger, L., Engelhardt, S., Roß, T., Clarkson, M. J., Gurusamy, K., Davidson, B. R., Maier-Hein, L., Riediger, C., Welsch, T., Weitz, J., & Speidel, S. (2019). *Generating Large Labeled Data Sets for Laparoscopic Image Processing Tasks Using Unpaired Image-to-Image Translation*. In: Shen, D., et al. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. *MICCAI 2019. Lecture Notes in Computer Science()*, vol 11768. Springer, Cham. https://doi.org/10.1007/978-3-030-32254-0_14
- Souipas, S., Nguyen, A., Laws, S. G., Davies, B. L., & Rodriguez y Baena, F. (2023). *SimPS-Net: Simultaneous Pose and Segmentation Network of Surgical Tools*. *IEEE Transactions on Medical Robotics and Bionics*, 5(3), 614. <https://doi.org/10.1109/TMRB.2023.3291022>
- Valbuena-Pardal R. (2022) *Sistema para la detección de instrumental quirúrgico en vídeos de cirugía laparoscópica empleando redes neuronales*. Universidad de Valladolid. <https://uvadoc.uva.es/bitstream/handle/10324/54177/TFG-I-2238.pdf?sequence=1>
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). *Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression*. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2019.00075>
- Zhang, Z., Chen, Y., Li, H., & Zhang, Q. (2021). *IA-CNN: A generalised interpretable convolutional neural network with attention mechanism*. *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, 2021, pp. 1-8. <https://doi.org/10.1109/IJCNN52387.2021.9533727>

