

**DETECCIÓN AUTOMÁTICA DE PÓLIPOS USANDO LA TRANSFORMADA
DENSE DE HOUGH EN SECUENCIAS COLONOSCÓPICAS**

LINA MARCELA RUIZ GARCÍA

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2020

**DETECCIÓN AUTOMÁTICA DE PÓLIPOS USANDO LA TRANSFORMADA
DENSE DE HOUGH EN SECUENCIAS COLONOSCÓPICAS**

LINA MARCELA RUIZ GARCÍA

**Una tesis presentada en cumplimiento de los requisitos para el grado de:
Ingeniera de Sistemas e Informática**

Director:

Fabio Martínez Carrillo

Doctor en Ingeniería de sistemas y computación

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2020

AGRADECIMIENTOS

El autor expresa su agradecimiento:

En primera instancia, a Dios y mi familia, que en todo momento me han apoyado en lo que anhelo, me han enseñado a luchar por mis sueños, a esforzarme y a lograr todo lo que me proponga. Siempre han estado ahí siendo mi mayor motivación en esta etapa tan importante, sin ellos esto no sería posible.

A mis padres, que son lo más importante que tengo, porque siempre con sus palabras y gestos de amor me han impulsado a dar lo mejor de mí. Son mi motor para ser una mejor persona y seguir creciendo en todos los sentidos. Espero poder recompensar todo aquello que me han dado.

A mi hermano, que me ha enseñado que sin interesar los problemas o circunstancias que se nos presenten siempre se debe seguir adelante con la frente en alto, que no importa si nos dicen que no podemos, o nadie cree en nosotros, no debemos dejar de luchar por lo que realmente queremos.

A mi director de proyecto, Fabio Martínez, por ser un excelente guía en este camino, por su paciencia y constancia. De él aprendí que la entrega, el esfuerzo, y perseverancia traerán gratas recompensas, que a veces sólo es necesario mirar al futuro y lo que se logrará para seguir luchando cada día. También le agradezco todo lo que me ha enseñado en este proceso, los conocimientos que me transmitió en el aula de clase, y principalmente por creer tanto en mí como en mis capacidades.

Al ingeniero Luis Guayacán que estuvo desde el inicio de este proyecto. Con sus pocas palabras, pero sabias, aportó a mi crecimiento, me apoyó, comprendió y me dio fortaleza cuando lo necesitaba. Estuvo presente en los momentos más importantes, dedicándome lo mejor de él.

A mis mejores amigos, Angélica y Sebastian, que han estado a mi lado desde hace muchos años, brindándome su apoyo en todo instante. Me dieron aliento, fuerzas para seguir adelante, y siempre han demostrado lo orgullosos que están de mí. Más que mis amigos son como mis hermanos, me alegra compartir este logro a su lado.

A todos mis compañeros y amigos de Ingeniería de Sistemas, especialmente a Alejandra, que con los años en cada salón de clase construimos una amistad inigualable. Empezamos este sueño juntas, y me da felicidad poder culminarlo de la misma manera. Siempre fue un apoyo incondicional, agradezco todos los consejos, motivación, y confianza que me brindó.

A mis amigos tanto de la universidad como del colegio, Alejandro, Martín, y demás personas que de alguna u otra forma aportaron en este proceso enseñándome grandes cosas.

Finalmente, a mis amigos del grupo de investigación *BIVL²AB*, Brayan, Oscar, Yesid, Franklin, Jefferson, Gustavo, y demás miembros, quienes me acompañaron en este proceso, brindándome su ayuda, y conocimiento en cualquier momento.

CONTENIDO

	pág.
INTRODUCCIÓN	10
1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA	17
2. OBJETIVOS	18
2.1. OBJETIVO GENERAL	18
2.2. OBJETIVOS ESPECÍFICOS	18
3. MÉTODO PROPUESTO	19
3.1. REPRESENTACIÓN DE PÓLIPOS NO PARAMÉTRICA	20
3.1.1. Aprendizaje de la plantilla del pólipos	20
3.1.2. Búsqueda y detección del pólipos	21
3.2. REPRESENTACIÓN LOCAL DE LOS PÓLIPOS	22
3.2.1. Características geométricas multi-escala	22
3.2.2. Características obtenidas por aprendizaje profundo	24
3.3. MAPAS DE ATENCIÓN	27
3.3.1. Máximo voto y similitud	28
3.3.2. Ubicación del pólipos a partir del detector de esquinas de Harris	29
4. DISEÑO EXPERIMENTAL	31
4.1. CONFIGURACIÓN DE LA EXTRACCIÓN DE CARACTERÍSTICAS	31
4.1.1. Características geométricas multi-escala	31
4.1.2. Características obtenidas por aprendizaje profundo	31
4.2. CONJUNTO DE DATOS	34

5. EVALUACIÓN Y RESULTADOS	36
6. CONCLUSIONES Y PERSPECTIVAS	48
BIBLIOGRAFÍA	51
ANEXOS	57

LISTA DE FIGURAS

	pág.
Figura 1. Esquema del método propuesto.	19
Figura 2. Representación de <i>Opponent Color</i>	24
Figura 3. Pre-procesamiento propuesto.	26
Figura 4. Mapas de atención	30
Figura 5. Mapas de características	33
Figura 6. Conjunto de datos	35
Figura 7. Detección del pólipo en diferentes fotogramas y su respectiva matriz acumulativa Γ_t	37
Figura 8. Operadores locales de búsqueda con el método de extracción de características obtenidas por aprendizaje profundo	39
Figura 9. Operadores locales de búsqueda con el método de extracción de características geométricas multi-escala	40
Figura 10. Extraer Θ regiones en cada mapa de atención Γ_t con la estrategia de características obtenidas por aprendizaje profundo	42
Figura 11. Extraer Θ regiones en cada mapa de atención Γ_t con características geométricas multi-escala	43
Figura 12. Detección de pólipos en las secuencias de colonoscopia	45
Figura 13. Resultados basados en fotogramas para las características obteni- das por aprendizaje profundo evaluando los conjuntos de datos <i>ASU-Mayo</i> y <i>CVC</i>	46
Figura 14. Resultados basados en fotogramas para las características geomé- tricas multi-escala evaluando los conjuntos de datos <i>ASU-Mayo</i> y <i>CVC</i> . . .	47

RESUMEN

TÍTULO: DETECCIÓN AUTOMÁTICA DE PÓLIPOS USANDO LA TRANSFORMADA DENSA DE HOUGH EN SECUENCIAS COLONOSCÓPICAS *

AUTOR: LINA MARCELA RUIZ GARCÍA **

PALABRAS CLAVE: CÁNCER DE COLON, DETECCIÓN DE PÓLIPOS, TRANSFORMADA DENSA DE HOUGH, CARACTERÍSTICAS OBTENIDAS POR APRENDIZAJE PROFUNDO.

DESCRIPCIÓN: Los pólipos son masas protuberantes que crecen a lo largo del tracto intestinal y se consideran los principales precursores del cáncer de colon. Su detección precisa en etapas tempranas representa una probabilidad de supervivencia del 93 % mientras que en otras etapas disminuye hasta un 8 %. La colonoscopia es el método actual más efectivo y ampliamente usado para detectar pólipos. Sin embargo, varios estudios han demostrado que durante este procedimiento se presenta una alta tasa de pérdida de pólipos reportando hasta un 26 %, debido a error humano, movimientos abruptos de la cámara, etapa de la enfermedad, estructuras intestinales similares a los mismos, entre otros. Las herramientas computacionales han surgido como una alternativa para apoyar la localización y detección automática de pólipos, pero muchos problemas permanecen abiertos debido a la alta variabilidad global y los cambios dramáticos tanto en forma como en aspecto característico de estas protuberancias, que tienen una forma no paramétrica y enormes variaciones entre los videos de colonoscopia. El presente trabajo introduce una estrategia computacional que produce mapas de atención visual con la ubicación más probable de los pólipos, para generar alarmas y apoyar la tarea de detección durante la colonoscopia. Siguiendo una representación de Hough densa, se elabora un modelo robusto del pólipo con tablas de características, que posteriormente se propagan a lo largo de una secuencia mediante un proceso de búsqueda y votación, para obtener mapas de atención visual con la ubicación más probable de masas anormales. Esta transformada integra características geométricas multi-escala y características profundas obtenidas de redes neuronales convolucionales pre-entrenadas. El método propuesto fue evaluado en secuencias de video de los conjuntos de datos *ASU-Mayo Clinic* y *CVC-Video Clinic*, reportando entre las cuatro regiones más probables una exactitud del 70 % en la detección de pólipos, mientras que con diez regiones 80 %.

* Trabajo de grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez, Ph.D.

ABSTRACT

TITLE: AUTOMATIC DETECTION OF POLYPS USING HOUGH'S DENSE TRANSFORM IN COLONOSCOPIC SEQUENCES. *

AUTHOR: LINA MARCELA RUIZ GARCÍA **

KEYWORDS: COLORECTAL CANCER, DETECTION OF POLYPS, HOUGH'S DENSE TRANSFORM, DEEP FEATURES.

DESCRIPTION: Polyps are protruding masses that grow along the intestinal tract and are considered to be the main precursors of colorectal cancer. Their accurate detection, in early stages, represents a survival probability up 93 %, while for other stages this probability decrease to 8 %. The colonoscopy is the current, most effective and widely used method to detect polyps. However, several studies have shown that during this procedure there is a high rate polyp loss detection up to 26 %, due to human error, sudden camera movements, stage of disease, polyp-like intestinal structures, among others. Computer tools have emerged as a robust alternative to support location and automatic detection of polyps, but many problems remain open due to the high global variability and dramatic changes in both the shape and appearance characteristic of polyps, having a non-parametric shape and huge variations among colonoscopy videos. The present work introduces a computational strategy that produces visual attention maps with the most likely location of the polyps, to generate alarms and to support the detection task during colonoscopy. Following a dense Hough representation, a polyp template is robustly modelled with indexes tables, which are thereafter propagated along video sequence through a search and voting process to obtain frame-visual attention maps with most probable location of abnormal masses. This transformation integrates multiscale geometrical features and dep features obtained from pre-trained convolutional neural network (CNN) architecture. The proposed method was evaluated on video sequences of the Asu-Mayo Clinic and CVC-Video Clinic datasets, reporting among the four most likely regions an accuracy of 70 % in detection, while ten regions yielded 80 %.

* Degree work

** Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez, PhD.

INTRODUCCIÓN

De acuerdo con *Global Cancer Observatory* (Globocan) ¹, el cáncer de colon es el tercer cáncer más agresivo a nivel mundial, reportando más de 1.5 millones de casos detectados y más de ochocientas mil muertes asociadas en 2018. En estudios más específicos, una dramática cifra de incidencia de cáncer de colon muestra, por ejemplo, que en América se prevé un crecimiento del 60 % para el año 2030, con un total de 396.000 nuevos casos por año ². Actualmente, diversos estudios han demostrado que la probabilidad de supervivencia del paciente (*SP-Survival Probability*) a los cinco años depende de la detección temprana, el tratamiento y seguimiento de la enfermedad, con un 93 % de SP si la detección se realiza en la primera etapa y un 8 % de SP si se realiza en la cuarta etapa de la enfermedad ^{3, 4}. Por lo tanto, el diagnóstico precoz del cáncer de colon resulta fundamental para llevar a cabo un tratamiento eficaz y exitoso.

La colonoscopia es considerada hoy en día como el recurso más efectivo para caracterizar y detectar el cáncer de colon, especialmente en etapas tempranas de la

¹ Globocan: <https://gco.iarc.fr/>

² FERLAY J Colombet M, Soerjomataram I Mathers C Parkin DM Piñeros M Znaor A Bray F. "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods doi: 10.1002/ijc.31937. [Epub ahead of print] PubMed PMID: 30350310". En: *Int J Cancer*. (2018 Oct 23).

³ PEREZ, Eduardo. *Gastroenterología*. McGraw Hill Mexico, 2012.

⁴ BASAVE, Horacio Noé López, *et al.* "Cáncer de colon y recto". En: *Manual de Oncología, 6e*. Ed. por HERRERA-GÓMEZ, Ángel; SILVA, Silvio A. Ñamendys y MENESES-GARCÍA, Abelardo. New York, NY: McGraw-Hill Education, 2018.

enfermedad ⁵. Este procedimiento clínico permite explorar el tracto intestinal y visualizar regiones anormales en los pliegues para evaluar, caracterizar e incluso eliminar neoplasias potencialmente malignas. Durante la colonoscopia, la evaluación más importante consiste en la detección de pólipos, los cuales son masas protuberantes anormales, que constituyen los principales biomarcadores del cáncer de colon ⁶. Sin embargo, la detección de pólipos a lo largo del tracto intestinal es una tarea muy compleja que requiere observaciones exhaustivas de expertos, las cuales toman alrededor de 20-30 minutos durante la rutina clínica ⁷. A pesar de esto, varios informes evidencian que en un procedimiento típico de colonoscopia se produce una tasa de detección errónea de pólipos entre un 4-12% ^{8, 9}. Esta tasa está principalmente asociada a la experiencia del médico, la preparación previa al procedimiento por parte del paciente e incluso a la etapa de la enfermedad. Aún peor, en recientes estudios clínicos se ha demostrado que la tasa de pérdida de estas protuberancias es de hasta un 25-26%, un factor dramático asociado principalmente a los pólipos

-
- ⁵ DAVILA, Raquel E.; RAJAN, Elizabeth y BARON, Todd H. "ASGE guideline: colorectal cancer screening and surveillance". En: *Gastrointestinal Endoscopy* 63.4 (2006), págs. 546 -557. DOI: <https://doi.org/10.1016/j.gie.2006.02.002>.
- ⁶ JESSICA COHAN, MD y MADHULIKA G VARMA, MD. "Intestino grueso". En: *Diagnóstico y tratamiento quirúrgicos, 14e*. Ed. por DOHERTY, Gerard M. New York, NY: McGraw-Hill Education, 2018.
- ⁷ KUPERIJ, N., *et al.* "Design of a user interface for intuitive colonoscope control". En: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2011, págs. 937-942.
- ⁸ TAJBAKHSI, N.; GURUDU, S. R. y LIANG, J. "Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information". En: *IEEE Transactions on Medical Imaging* 35.2 (2016), págs. 630-644.
- ⁹ WANG, Y., *et al.* "Part-Based Multiderivative Edge Cross-Sectional Profiles for Polyp Detection in Colonoscopy". En: *IEEE Journal of Biomedical and Health Informatics* 18.4 (2014), págs. 1379-1389.

diminutos ($\sim [5 - 9]$ mm) ^{10, 11, 12}, los cuales están presentes en las primeras etapas de la enfermedad, y su identificación constituye una oportunidad inestimable para lograr la eficacia del tratamiento, hecho que se relaciona con una mayor probabilidad de supervivencia del paciente ^{13, 14, 12}. En contraste, nuevos estudios sugieren que algunos pacientes pueden desarrollar cáncer de colon dentro de 3 a 5 años después de haber realizado esta rutina clínica ^{15, 16}.

Por lo tanto, los sistemas computacionales para apoyar los procedimientos de colonoscopia, se han convertido en herramientas fundamentales para detectar, segmen-

¹⁰ ANGERMANN, Quentin; HISTACE, Aymeric y ROMAIN, Olivier. "Active learning for real time detection of polyps in videocolonoscopy". En: *Procedia Computer Science* 90 (2016), págs. 182-187.

¹¹ BRESSLER, Brian, *et al.* "Rates of New or Missed Colorectal Cancers After Colonoscopy and Their Risk Factors: A Population-Based Analysis". En: *Gastroenterology* 132.1 (2007), págs. 96-102. DOI: <https://doi.org/10.1053/j.gastro.2006.10.027>.

¹² VAN RIJN, Jeroen C, *et al.* "Polyp miss rate determined by tandem colonoscopy: a systematic review". En: *The American journal of gastroenterology* 101.2 (2006), pág. 343.

¹³ POINTET, A-L y TAIEB, J. "Autoevaluaciones del artículo: Cáncer de colon". En: (2016).

¹⁴ BERNAL, J., *et al.* "Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge". En: *IEEE Transactions on Medical Imaging* 36.6 (2017), págs. 1231-1249.

¹⁵ CLERCQ, Chantal MC le , *et al.* "Postcolonoscopy colorectal cancers are preventable: a population-based study". En: *Gut* 63.6 (2014), págs. 957-963.

¹⁶ RABENECK, Linda y PASZAT, Lawrence F. "Circumstances in which colonoscopy misses cancer". En: *Frontline gastroenterology* 1.1 (2010), págs. 52-58.

tar, clasificar y, en general, caracterizar mejor los pólipos ^{17, 18, 19}. Sin embargo, con respecto al modelado de pólipos, la caracterización de tales masas es un desafío, debido a su amplia variabilidad en cuanto a forma, color y patrones de textura. Además de esto, cabe considerar la estrecha similitud que tienen con la pared del tracto intestinal en sus regiones vecinas, y que junto con los retos inherentes al procedimiento clínico como la resolución de la cámara del colonoscopio, los movimientos bruscos de la misma durante la exploración y la fuerte variación de un fotograma a otro, hacen que la tarea de reconocimiento de estos biomarcadores sea considerablemente difícil, incluso para los gastroenterólogos expertos ³.

En cuanto a la detección (semi)-automática de pólipos, los enfoques propuestos pueden clasificarse según las características de forma, textura o color que los modelan ²⁰. Por ejemplo, algunos trabajos han representado la forma de los pólipos a partir del análisis de bordes y del valle de intensidad ²¹ o usando la transformada de

¹⁷ TAJBAKHS, Nima; GURUDU, Suryakanth R y LIANG, Jianming. "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks". En: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2015, págs. 79-83.

¹⁸ GANZ, Melanie; YANG, Xiaoyun y SLABAUGH, Greg. "Automatic segmentation of polyps in colonoscopic narrow-band imaging data". En: *IEEE Transactions on Biomedical Engineering* 59.8 (2012), págs. 2144-2151.

¹⁹ STEHLE, Thomas, *et al.* "Classification of Colon Polyps in NBI Endoscopy Using Vascularization Features". En: *Medical Imaging 2009: Computer-Aided Diagnosis*. Ed. por KARSSEMEIJER, N. y GIGER, M. L. Vol. 7260. Orlando, USA: SPIE, 2009.

²⁰ BERNAL, Jorge, *et al.* "Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge". En: *IEEE transactions on medical imaging* 36.6 (2017), págs. 1231-1249.

²¹ BERNAL, Jorge; SÁNCHEZ, Javier y VILARINO, Fernando. "Towards automatic polyp detection with a polyp appearance model". En: *Pattern Recognition* 45.9 (2012), págs. 3166-3182.

Hough de segundo orden ²². Sin embargo, estos métodos generan un alto número de falsos positivos debido a las limitaciones en el modelado de la variabilidad de estas protuberancias y la estrecha similitud con las regiones circundantes. Otros enfoques han dedicado esfuerzos para caracterizar los pólipos a partir de sus patrones de textura y color, planteando estrategias basadas en *Local Binary Patterns* (LBP) ²³, *Color Wavelet Covariance* (CWC) ²⁴ y *Histogram of Oriented Gradients* (HOG) ²⁵. Estos métodos son propensos al error cuando hay cambios bruscos de iluminación y movimientos repentinos de la cámara. Algunas aproximaciones adicionales involucran la geometría y descriptores de color del pólipo, pero sigue siendo limitado para tratar su variabilidad ²⁶.

Trabajos más recientes han integrado representaciones de aprendizaje profundo para tratar la caracterización del pólipo. Por ejemplo, la arquitectura *Faster Region with Convolutional Neural Network* (R-CNN), que incluye dos módulos para caracterizar

²² TARIK, Ghannam, *et al.* "Polyps's region of interest detection in colonoscopy images by using clustering segmentation and region growing". En: *Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on*. IEEE. 2016, págs. 455-459.

²³ AMELING, Stefan, *et al.* "Texture-Based Polyp Detection in Colonoscopy". En: *Bildverarbeitung für die Medizin 2009*. Ed. por MEINZER, Hans-Peter, *et al.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, págs. 346-350.

²⁴ IAKOVIDIS, Dimitrios K, *et al.* "A comparative study of texture features for the discrimination of gastric polyps in endoscopic video". En: *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*. IEEE. 2005, págs. 575-580.

²⁵ IWAHORI, Yuji, *et al.* "Automatic detection of polyp using Hessian Filter and HOG features". En: *Procedia computer science* 60 (2015), págs. 730-739.

²⁶ TAJBAKHS, Nima; GURUDU, Suryakanth R y LIANG, Jianming. "Automated polyp detection in colonoscopy videos using shape and context information". En: *IEEE transactions on medical imaging* 35.2 (2016), págs. 630-644.

y localizar objetos, ha sido adaptada a este problema de caracterización ²⁷. También, Zheng *et al.* adaptó la *U-Net* que junto con el flujo óptico obtiene información temporal y logra detecciones adecuadas de los pólipos ²⁸. En ²⁹ se implementa la red *RYCO*, una extensión de *You Only Look Once* (YOLO), que adapta un filtro de correlación discriminatorio para obtener información temporal y refinar la detección. Estos enfoques han demostrado ser eficaces modelando la variación de los pólipos basado en representaciones visuales convolucionales y no lineales. Sin embargo, estas arquitecturas requieren grandes conjuntos de datos para entrenamiento, siendo una limitación principal en el dominio clínico. Por consiguiente, la tarea de cuantificar, identificar y caracterizar automáticamente estos biomarcadores sigue siendo un problema abierto, especialmente para su utilización en escenarios en tiempo real con grandes variaciones en su apariencia, y en general, en escenas grabadas en video-colonoscopias.

Este trabajo introduce una estrategia computacional que produce mapas de atención visual, con las ubicaciones más probables de los pólipos, para apoyar la tarea de detección de los mismos durante el procedimiento de colonoscopia. Para ello, cada fotograma de una secuencia se representa por un conjunto de características locales. En este método se utilizaron dos enfoques: 1) las características denominadas geométricas multi-escala, que se extraen de manera manual junto a un análisis multi-escala calculando diversos filtros Gaussianos incrementales, y 2) las caracte-

²⁷ MO, X., *et al.* "An Efficient Approach for Polyps Detection in Endoscopic Videos Based on Faster R-CNN". En: *2018 24th International Conference on Pattern Recognition (ICPR)*. 2018, págs. 3929-3934. DOI: 10.1109/ICPR.2018.8545174.

²⁸ ZHENG, H., *et al.* "Polyp Tracking in Video Colonoscopy Using Optical Flow With an On-The-Fly Trained CNN". En: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019, págs. 79-82. DOI: 10.1109/ISBI.2019.8759180.

²⁹ ZHANG, Ruikai, *et al.* "Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker". En: *Pattern Recognition* 83 (2018), págs. 209 -219. DOI: <https://doi.org/10.1016/j.patcog.2018.05.026>.

rísticas obtenidas por aprendizaje profundo, extraídas de una red neuronal convolucional profunda (CNN - *Convolutional Neural Network*) pre-entrenada, obteniendo así una descripción enriquecida y general de las respuestas de la primera capa de la CNN. Cada mapa profundo codifica características particulares de los pólipos y permite una representación visual no lineal de los mismos. A partir de una plantilla del pólipo, extraída del primer fotograma, se realiza una codificación densa (por píxel) siguiendo el esquema de la transformada densa de Hough (TDH)³⁰, esta codificación permite una representación no paramétrica de la forma de los pólipos. En cada fotograma, los mapas de atención visual contienen la ubicación más probable del pólipo y se obtienen mediante un proceso de búsqueda en tablas representativas que votan por una posible ubicación del biomarcador. En los diferentes mapas de atención, se utilizaron varias estrategias para localizar más regiones potenciales con masas protuberantes, como por ejemplo, la similitud de coseno y detector de esquinas. Las ubicaciones más probables se proyectan al espacio de la imagen a lo largo de la secuencia de colonoscopia.

³⁰ MANZANERA, Antoine. "Dense Hough transforms on gray level images using multi-scale derivatives". En: *SIXIEME WORKSHOP AMINA 2012 Applications Médicales de l'Informatique: Nouvelles Approches*. 2012.

1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

Los pólipos son el principal biomarcador del cáncer de colon, siendo vital el diagnóstico precoz de esta enfermedad para la supervivencia del paciente. Según reportes de la literatura, el 90 % de los casos pueden ser tratados exitosamente si se diagnostican en las primeras etapas de la enfermedad. Dichos pólipos, durante la colonoscopia, son caracterizados como masas protuberantes que crecen en la pared interna del intestino.

A pesar de la importancia de su detección temprana, se estima una tasa de pérdida del 26 % de pólipos durante la rutina clínica. En general, esta tasa de pérdida está relacionada con los pequeños tamaños, reportados en etapas tempranas, su morfología y apariencia altamente correlacionada con las paredes intestinales. Cabe destacar que el proceso de análisis y detección en las colonoscopias es realizado únicamente por inspección visual de gastroenterólogos expertos. Visualmente la caracterización de los pólipos es compleja debido a las altas variaciones en cuanto a forma, tamaño, color y textura. Además, durante el examen, los pólipos pueden presentarse en diferentes escalas que se distorsionan teniendo en cuenta los cambios abruptos de la cámara del colonoscopio.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

- Codificar características visuales colonoscópicas usando la transformada densa de Hough para la detección automática de pólipos.

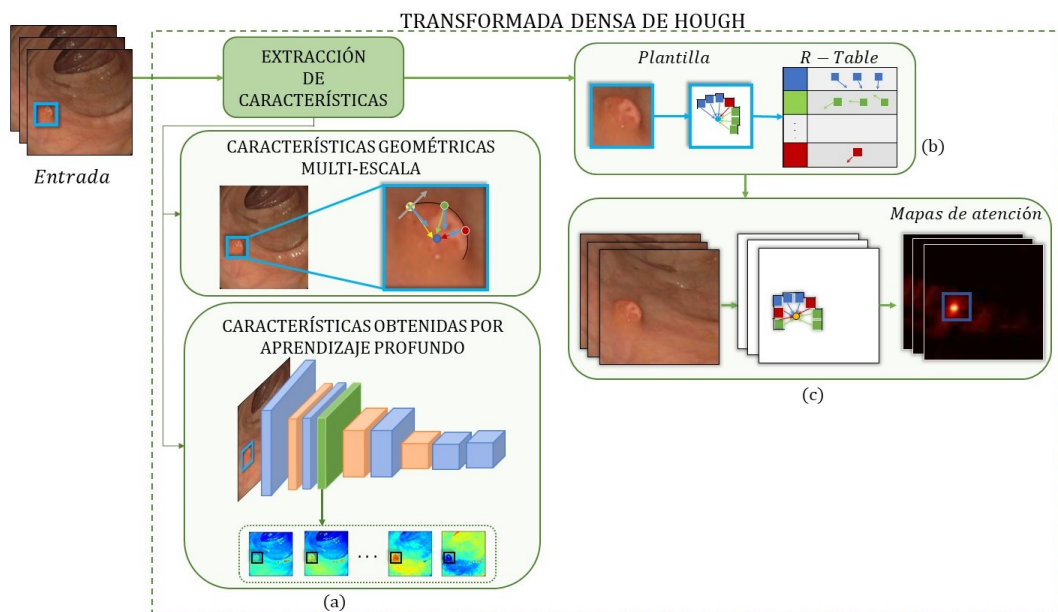
2.2. OBJETIVOS ESPECÍFICOS

- Procesar un conjunto de secuencias colonoscópicas que permitan resaltar características relacionadas con los pólipos.
- Implementar estrategias de extracción de características locales en las secuencias colonoscópicas.
- Codificar características visuales en cada fotograma usando la transformada densa de Hough.
- Desarrollar una estrategia para detección de regiones de interés en mapas acumulados de votación.
- Validar el método usando diferentes métricas de localización y detección de regiones con potencial presencia de pólipos.

3. MÉTODO PROPUESTO

El presente trabajo introduce un enfoque novedoso que consiste en extraer una muestra del pólipo en el primer fotograma y construir mapas de atención de pólipos para cada fotograma de la secuencia, con el fin de soportar la detección y caracterización de los mismos durante el procedimiento de la colonoscopia. De la muestra del pólipo se extraen características a partir de un conjunto de características locales y se codifican con tablas acumulativas. Luego, se lleva a cabo un proceso de votación, teniendo como base aquellas características extraídas, sobre la posición de cada píxel para determinar la ubicación más probable de las masas anormales. El esquema general del método propuesto se presenta en la figura 1.

Figura 1. A partir de un video de entrada, se realiza un proceso de extracción de dos tipos de características que son aprendidas por la TDH, generando tablas que almacenan la información y mapas de votación que contienen la probabilidad de la ubicación del pólipo en cada fotograma. (a) Proceso de extracción de características. (b) Codificación de la plantilla del pólipo en la *R-Table*. (c) Búsqueda y detección del pólipo a partir de los mapas de atención visual obtenidos para cada fotograma.



3.1. REPRESENTACIÓN DE PÓLIPOS NO PARAMÉTRICA

Los pólipos se observan como formas no paramétricas con alta variabilidad geométrica y de apariencia intra e inter paciente. Por lo tanto, un modelado local de estas masas es exigente para hacer frente a tales variaciones. En este trabajo, el principio de la transformada de Hough fue adaptado localmente para modelar dichos biomarcadores como representaciones no paramétricas, mediante la implementación de la transformada densa de Hough (TDH), un esquema de votación de características multidimensional que opera a nivel de píxel y permite construir representaciones densas flexibles e independientes de estructuras geométricas o de apariencia ³⁰. En particular, la región del pólipo se codifica en una plantilla que se mapeará a través de una secuencia de video, destacando las regiones espaciales con mayor probabilidad de coincidencia de estas masas protuberantes, es decir, mapas de atención de pólipos, que resumen las respuestas en el espacio multidimensional de Hough. Este proceso se organiza en dos fases denominadas: el aprendizaje de la plantilla del pólipo, y la búsqueda y detección de dicha plantilla en fotogramas consecutivos, que se describen en las siguientes subsecciones.

3.1.1. Aprendizaje de la plantilla del pólipo Una representación visual densa de los pólipos se logra aprendiendo una plantilla en el primer fotograma de la secuencia. En primer lugar, para cualquier secuencia de colonoscopia, se obtiene una representación en cada píxel \mathbf{x} en una ubicación determinada $(m, k)^T$ codificadas a lo largo de los ejes M y K , mediante un vector de características $\mathbf{x} \in \mathbb{R}^N$. Cada fotograma $I_t \in \mathbb{R}^{(w \times h \times N)}$ es entonces una matriz multidimensional con N características y tamaño de $(w \times h)$, siendo t el número del fotograma. Una plantilla particular T_r es una sub-matriz de la imagen, que en el primer fotograma se selecciona manualmente para delimitar una muestra del pólipo. Esta sub-región T_r , con un tamaño $(w_T \times h_T)$ que varía dependiendo de las dimensiones de la protuberancia, se codifi-

ca para representar la complejidad visual del biomarcador. Como en la codificación de Hough, una nueva representación se obtiene calculando las distancias de cada píxel con el centro de la plantilla, y son indexadas por los contenedores de características definidas ($\phi(x_i)$), específicamente, para un píxel de característica particular x_i en la posición (m, k, i) se calcula una distancia ($\delta_M = [\frac{w_T}{2} - m], \delta_K = [\frac{h_T}{2} - k]$), siendo $(\frac{w_T}{2}, \frac{h_T}{2})$ el centro de la plantilla T_r . Cada una de las distancias obtenidas es ponderada de acuerdo a un valor (ω) que se establece según la importancia de la característica, con esta información se construye y almacena la tupla $(\delta_M, \delta_K, \omega)$. Estas características son indexadas de acuerdo al dominio de la misma, formando la tabla de representación (R-Table - *Representation Table*), definida como $RT(\phi(\mathbf{x}))$, que constituye un descriptor de pólipos efectivo. Este proceso se ilustra en la figura 1.(b) y es representado en el Algoritmo 1 en la función definida como R-Table().

3.1.2. Búsqueda y detección del pólipo Los mapas de atención Γ_t se obtienen aquí en cada fotograma de la secuencia, proyectando la tabla de representación $RT(\phi(\mathbf{x}))$ aprendida de la plantilla (T_r), al resto del video de colonoscopia. Estos mapas Γ_t son matrices con la misma dimensión de los fotogramas ($w \times h$) que almacenan la ubicación más probable de los pólipos. Para asignar la probabilidad del pólipo en cada píxel \mathbf{x} (con ubicación (m, k)), cada valor de la característica x_i señala un contenedor $\phi(\mathbf{x})$ en la correspondiente $RT(\phi(\mathbf{x}))$. El contenedor $\phi(x_i)$ seleccionado almacena un conjunto de distancias aprendidas $(\delta_M + m, \delta_K + k)$ que se proyectan desde la ubicación de los píxeles (m, k) . Cada una de las distancias proyectadas realiza un voto acumulativo, con una importancia ponderada particular (ω), en el píxel con la posición $(\delta_M + m, \delta_K + k)$. En tal caso, si la mayoría de los píxeles votan por la misma posición, este píxel correspondiente tendrá mayor importancia, es decir, será más probable que esta ubicación corresponda al centro del pólipo. El proceso se lleva a cabo en todo el espacio de características y en cada fotograma de la secuencia. La figura 1.(c) ilustra tal proceso para formar los mapas de atención de

colonoscopia, y se resume en el Algoritmo 1 en la función definida como Gamma().

Algoritmo 1. TDH. Aprendizaje de la plantilla y creación de la matriz Γ_t

Función R-Table():

```

 $x \in \{x_1, \dots, x_i, \dots, x_N\}$ 
para  $x \in T_r$  hacer
    si  $x_i(m, k) > 0$  entonces
         $RT(\phi(x_i)) := RT(\phi(x_i)) \cup (\delta_M, \delta_K, \omega)$ 
    fin
fin

```

devolver

Función Gamma():

```

 $x \in \{x_1, \dots, x_i, \dots, x_N\}$ 
para  $x \in I_t$  hacer
    si  $x_i(m, k) > 0$  entonces
        calcular( $\phi(x_i)$ )
        para todo  $(\delta_M, \delta_K, \omega) \in RT(\phi(x_i))$  hacer
             $\Gamma_t(\delta_M + m, \delta_K + k) \leftarrow \Gamma_t(\delta_M + m, \delta_K + k) + \omega$ 
        fin
    fin
fin

```

devolver

3.2. REPRESENTACIÓN LOCAL DE LOS PÓLIPOS

3.2.1. Características geométricas multi-escala Cada fotograma I_t es caracterizado usando características geométricas denominadas: orientación del gradiente (α), curvatura (κ) y el canal de color (O_1). Para ello, se implementa un análisis multi-escala mediante el cálculo de varios filtros Gaussianos incrementales (σ). Los gradientes se calculan sobre cada escala (σ) de la imagen para obtener la magni-

tud de gradiente $\|\nabla I_t\| = \sqrt{I_{t_M}^2 + I_{t_K}^2}$ y el ángulo $\alpha = \arctan \frac{I_{t_K}}{I_{t_M}}$, usando derivadas de primer orden. La curvatura κ es estimada para cada píxel usando derivadas de segundo orden, expresada como:

$$\kappa = \frac{I_{t_{MM}}I_{t_K}^2 - 2I_{t_{MK}}I_{t_M}I_{t_K} + I_{t_{KK}}I_{t_M}^2}{\|\nabla I_t\|^3}$$

Finalmente, el canal de color O_1 se obtiene de la representación de *Opponent Color*³¹, implementado para mitigar los problemas de iluminación que tienen los pólipos a causa de los movimientos bruscos de la cámara durante el procedimiento de colonoscopia. Esta representación de color es normalizada a partir de las componentes RGB (del inglés *Red, Green, Blue*) del fotograma expresada como:

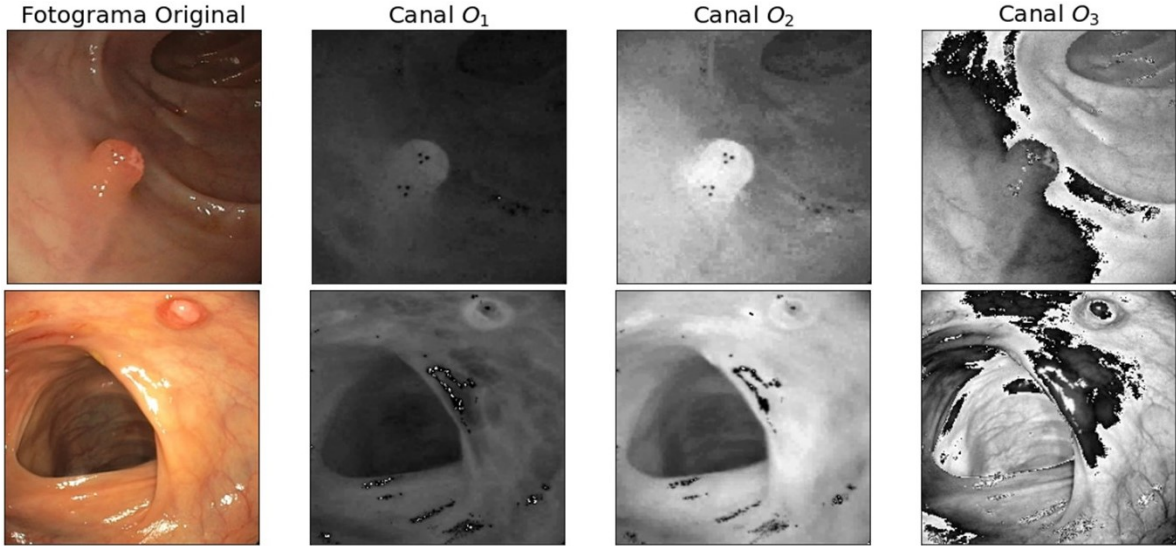
$$O_1 : \frac{R - G}{\sqrt{2}} \quad O_2 : \frac{R + G - 2B}{\sqrt{6}} \quad O_3 : \frac{R + G + B}{\sqrt{3}}$$

Debido a que la resta en los canales O_1 y O_2 , resultan en una relativa variación de desplazamiento con respecto a la luz, tal representación de color, permite una mejor visualización del pólipo en comparación al resto del fotograma. En este trabajo se implementó dicha estrategia usando sólo el canal O_1 . En la figura 2 se ilustra un ejemplo de la representación de este espacio de color.

Por lo tanto, una forma no paramétrica como la del pólipo puede ser representada localmente mediante el cálculo de ángulos, curvaturas y resaltando sus propiedades de color $\mathbf{x} = \{\alpha_{(\sigma_1)}, \kappa_{(\sigma_1)}, O_{1(\sigma_1)}, \alpha_{(\sigma_2)}, \kappa_{(\sigma_2)}, O_{1(\sigma_2)}, \alpha_{(\sigma_3)}, \kappa_{(\sigma_3)}, O_{1(\sigma_3)}\}$ para cada escala gaussiana (σ_l) cuya norma del gradiente es mayor a cero. Los ángulos (α) están definidos en un rango de $[-90, 90]$, mientras que (κ) está definida en un rango de $[-1, 1]$. En este caso, los índices de la *R-Table* ($RT(\phi(\mathbf{x}))$) se representan mediante

³¹ VAN DE SANDE, Koen; GEVERS, Theo y SNOEK, Cees. "Evaluating color descriptors for object and scene recognition". En: *IEEE transactions on pattern analysis and machine intelligence* 32.9 (2010), págs. 1582-1596.

Figura 2. Representación de *Opponent Color*. Cada fila muestra un ejemplo de dicha representación para diferentes fotogramas en cada uno de los canales (O_1, O_2, O_3).



el cálculo de cada primitiva (α, κ) , y en el caso del color O_1 , por el valor de intensidad en cada píxel. Finalmente, el peso acumulativo (ω) se define a partir de la norma del gradiente $\|\nabla I_t\|$ para el ángulo (α) , la norma de Frobenius de la matriz Hessiana definida como: $\|H_{I_t}\|_F = \sqrt{I_{tMM}^2 + 2I_{tMK}^2 + I_{tKK}^2}$ para la curvatura (κ) mientras que para el canal de color O_1 se estableció como 1.

3.2.2. Características obtenidas por aprendizaje profundo El modelado de los pólipos es una tarea difícil que debe enfrentarse a una representación muy variable en las colonoscopias. Además, los reflejos especulares, los movimientos bruscos y otros artefactos aumentan la complejidad para recuperar estos biomarcadores de las observaciones de apariencia. En la literatura han surgido representaciones de aprendizaje profundo para modelar entidades visuales complejas, pero que requieren grandes cantidades de muestras de entrenamiento etiquetadas para lograr una solución eficaz de la tarea. De hecho, existen varios trabajos que tratan sobre el

modelado de pólipos a partir de redes profundas ³². Sin embargo, en lo que respecta a dicho modelado, y teniendo en cuenta la variabilidad de estas masas, existen pocos conjuntos de entrenamiento disponibles para desarrollar esta tarea y, por otra parte, la anotación manual de los pólipos es una tarea sumamente exigente para los expertos.

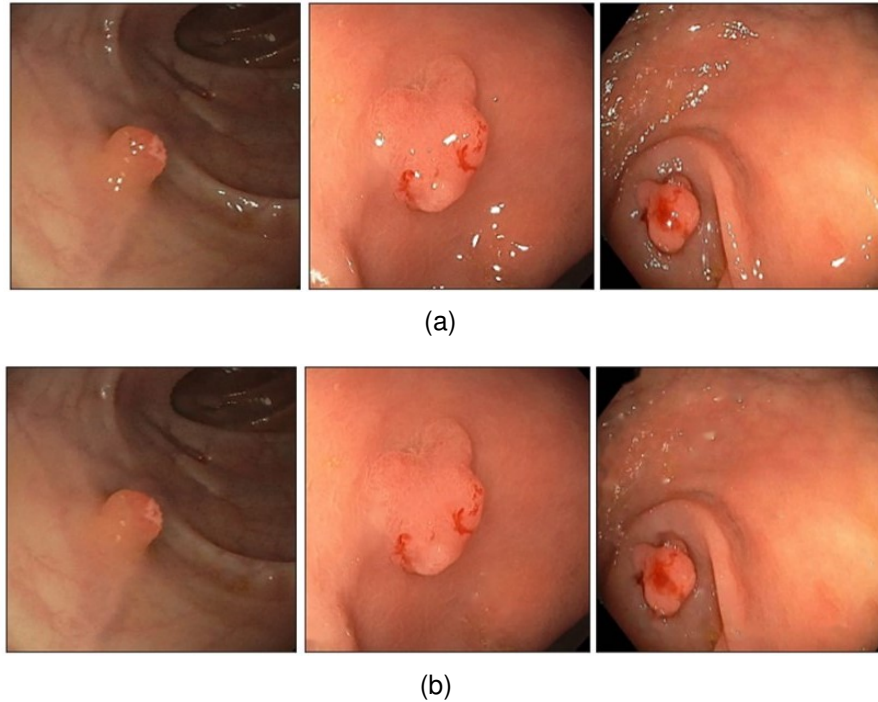
Por lo tanto, en este trabajo cada fotograma de colonoscopia I_t está plenamente representado mediante el uso de características obtenidas por aprendizaje profundo de redes convolucionales pre-entrenadas CNN's (del inglés *Convolutional Neural Networks*) que garantizan una representación visual adecuada. Cada fotograma se caracteriza por bloques de activación, aprendidos en representaciones de bajo nivel que consideran particiones no lineales del espacio. Se espera que la descomposición convolucional aprendida con filtros no lineales proporcione una mejor representación del pólipo y su variabilidad. En este caso, las características extraídas para cada píxel $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$ representan la respuesta de N filtros convolucionales en una capa primaria.

Para calcular las características obtenidas por aprendizaje profundo \mathbf{x} en cada fotograma de colonoscopia I_t se realiza en primer lugar un esquema de pre-procesamiento para eliminar los reflejos especulares que aparecen como regiones brillantes a lo largo del tracto intestinal, debido a la incidencia de la luz del colonoscopio sobre el mismo. Este proceso comienza transformando la representación de color de cada fotograma de RGB en HSV (del inglés *Hue, Saturation, Value*), y umbralizando aquellas regiones que tienen alto brillo (V) y baja saturación (S). Estas regiones segmentadas crecen mediante la implementación de operaciones morfológicas, específicamente la de dilatación. Finalmente, se aplica una estrategia de corrección

³² ALAGAPPAN, Muthuraman, *et al.* "Artificial intelligence in gastrointestinal endoscopy: The future is almost here". En: *World journal of gastrointestinal endoscopy* 10.10 (2018), pág. 239.

sobre el fotograma original ³³. Los resultados obtenidos de este pre-procesamiento se presentan en la figura 3.

Figura 3. Pre-procesamiento propuesto. (a) Fotogramas originales. (b) Respective fotogramas sin reflejos especulares.



Una vez que cada fotograma se filtra, los rasgos profundos convolucionales se obtienen mapeando los diversos fotogramas en la primera capa de una arquitectura conocida y previamente entrenada. Esta primera capa descompone la información visual en características de bajo nivel mediante el cálculo progresivo de transformaciones lineales seguidas de contracciones no lineales, proyectando la información sobre un conjunto de N filtros aprendidos $\Psi^j = \{\psi_1^j, \psi_2^j, \dots, \psi_N^j\}$ en una determinada capa j . Por lo tanto, cada fotograma I_t se filtra mediante un conjunto particular

³³ BERTALMIO, Marcelo; BERTOZZI, Andrea L y SAPIRO, Guillermo. "Navier-stokes, fluid dynamics, and image and video inpainting". En: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. IEEE. 2001, págs. I-I.

Ψ^j , obtenido de una representación de características convolucionales:

$$\mathbf{x} = \sum_{n=1 \dots N} \mathbf{I}_t * \psi_n^j$$

con x_i como la característica que resulta de convolucionar con respecto al filtro Ψ_i^j . Entonces, $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$ es el conjunto de características obtenidas por aprendizaje profundo que representan cada fotograma t . En este caso, el valor de intensidad presentado por cada activación representa el índice que es almacenado en $RT(\phi(\mathbf{x}))$, por lo tanto, no hay un rango de variación predeterminado para todos los casos, ya que esto implica una pérdida de información. Por otra parte, el peso (ω) se estableció como 1 para todas las activaciones. Estas características convolucionales preservan la estructura espacial y son capaces de representar patrones complejos de textura tanto locales como regionales que pueden ser eficaces para diferenciar entre los pólipos y los pliegues del tracto intestinal en cada fotograma.

3.3. MAPAS DE ATENCIÓN

Los mapas de atención de colonoscopia Γ_t , en cada fotograma I_t , representan la probabilidad sobre cada píxel de ser el centroide de la región del pólipo en particular. Estos mapas de atención se implementan a partir de la tabla de representación $RT(\phi(\mathbf{x}))$, del pólipo aprendido (T_r), en el espacio de características extraídas. Gracias a la representación de las características obtenidas, estos mapas de atención permiten recuperar eficazmente las regiones sospechosas, donde los gastroenterólogos deberían prestar atención para identificar masas anormales. No obstante, los pliegues del tracto intestinal y los artefactos de colonoscopia pueden producir un número considerable de regiones máximas en los mapas de atención Γ_t , es decir, los votos presentes en cada mapa se encuentran distribuidos tanto en el pólipo como en

dichos artefactos, obteniendo votaciones dispersas. Por lo tanto, se consideraron diferentes operadores locales de búsqueda de máximos para seleccionar las regiones de pólipos más probables sobre Γ_t . La ubicación detectada en Γ_t es propagada al fotograma original para definir potencialmente regiones como herramienta de apoyo adicional en la tarea de detección de estos biomarcadores. Se exploraron dos estrategias para seleccionar las regiones de máxima probabilidad, en la figura 4 se ilustra el proceso de cada una y se describen en las siguientes subsecciones.

3.3.1. Máximo voto y similitud Se define un primer operador máximo no lineal sobre el mapa de atención para seleccionar las regiones de pólipos más probables. En tal sentido, se implementan dos reglas para seleccionar estas regiones. Inicialmente, una región con las mismas dimensiones de la plantilla se superpone sobre el píxel de máximo voto (\max_{Γ_t}), esta selección representa la región más probable en el mapa de atención Γ_t . Todos los píxeles que están dentro de esta primera región se cancelan (se ponen en cero) para permitir la búsqueda de un segundo máximo en una ubicación diferente del mapa. Este proceso se realiza iterativamente obteniendo un conjunto de puntos (\max_{Γ_t}).

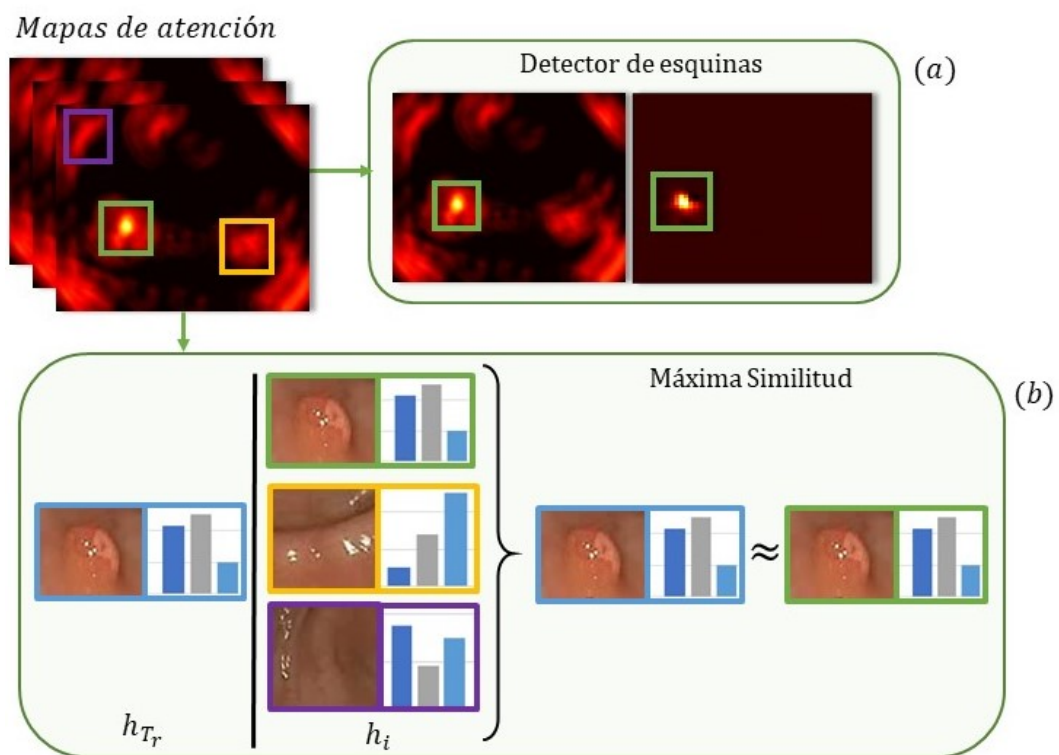
Por lo tanto, se introduce una regla para reforzar la selección de la ubicación máxima de los píxeles y evitar las regiones de falsos positivos relacionados principalmente con los pliegues del tracto intestinal. En este caso, la similitud en cuanto a la apariencia se mide calculando el histograma (\mathbf{h}_{T_r}) de la plantilla aprendida (T_r) y el histograma de la región (\mathbf{h}_i) extraída de Γ_t , que delimita la votación máxima escogida. Esta comparación se logra calculando la métrica de similitud del coseno, definida como: $\max_{T_r(\mathbf{h}_i)} \cos(\mathbf{h}_{T_r}, \mathbf{h}_i)$, sobre los histogramas de las regiones, lo cual permite ponderarlas y reordenarlas. Aquella que tenga una mayor similitud es la que se selecciona, y los píxeles centrales son propagados al fotograma original para marcar las regiones en donde es más probable encontrar pólipos.

3.3.2. Ubicación del pólipo a partir del detector de esquinas de Harris Una segunda estrategia considerada fue seleccionar un píxel de voto asociado a un máximo local, con bordes relevantes uniformes y dominantes alrededor del punto, es decir, una esquina en el mapa de atención $\Gamma_t(m, k)$. Siguiendo esta hipótesis, se implementó un detector de esquinas para analizar la importancia de la distribución local de cada punto a lo largo del eje espacial. Entonces, la región que delimita cada punto es representada mediante un tensor estructural codificado con la matriz Hessiana, expresada como:

$$H = \sum_{(m,k)} \begin{bmatrix} \Gamma_{t_M}^2 & \Gamma_{t_M} \Gamma_{t_K} \\ \Gamma_{t_M} \Gamma_{t_K} & \Gamma_{t_K}^2 \end{bmatrix}$$

Esta matriz codifica la información de borde a lo largo del eje espacial y la respectiva descomposición de los valores propios (λ_1, λ_2) permite determinar la importancia del píxel de votación, como: $\lambda_1 \lambda_2 - K(\lambda_1 + \lambda_2)^2$, con K como un valor constante. En este caso, sólo si ambos valores propios son relativamente grandes, el punto es considerado como la ubicación relevante del pólipo.

Figura 4. Operadores locales de búsqueda de máximos para seleccionar las regiones de pólipos más probables en cada mapa de atención Γ_t . (a) Ubicación del pólipo a partir del detector de esquinas de Harris. (b) Máxima similitud con la métrica de coseno.



4. DISEÑO EXPERIMENTAL

4.1. CONFIGURACIÓN DE LA EXTRACCIÓN DE CARACTERÍSTICAS

4.1.1. Características geométricas multi-escala En este caso se implementó, una representación densa a diferentes escalas, obtenidas mediante la convolución de cada fotograma de las secuencias colonoscópicas con un conjunto de filtros Gaussianos (σ_l). La mejor configuración se logró para un total de tres escalas $\sigma_l = \{3, 5, 7\}$. El número de tablas representativas (*R-Tables*) aprendidas es igual al número de características geométricas multi-escala extraídas, las cuales están representadas por el conjunto $\mathbf{x} = \{\alpha_{(\sigma_1)}, \kappa_{(\sigma_1)}, O_{1(\sigma_1)}, \alpha_{(\sigma_2)}, \kappa_{(\sigma_2)}, O_{1(\sigma_2)}, \alpha_{(\sigma_3)}, \kappa_{(\sigma_3)}, O_{1(\sigma_3)}\}$.

4.1.2. Características obtenidas por aprendizaje profundo La representación local y profunda de los píxeles de colonoscopia resulta fundamental para caracterizar los pólipos, codificadas en la tabla representativa $RT(\phi(\mathbf{x}))$. El número de *R-Tables* aprendidas es proporcional al número de características obtenidas por aprendizaje profundo extraídas en una determinada capa de la CNN. En este trabajo, el conjunto de características obtenidas por aprendizaje profundo se obtuvo de modelos pre-entrenados, utilizando las primeras capas, que han demostrado un rendimiento satisfactorio en las tareas clásicas de clasificación en imágenes. Cada una de las arquitecturas seleccionadas se describe a continuación:

- **Residual Network Model (ResNet):** es una exitosa red, ganadora de ILSVRC (del inglés, *ImageNet Large Scale Visual Recognition Challenge*)³⁴, un desafío de clasificación de imágenes naturales y detección de objetos, cuyo conjunto de datos cuenta con más de un millón de imágenes distribuidas en 1.000 cla-

³⁴ ILSVRC: <http://www.image-net.org/>

ses . La principal contribución de esta red es la implementación de bloques residuales que se ocupan del desvanecimiento del gradiente y permite representaciones visuales más profundas. Estos bloques residuales reciben durante la entrada conexiones de salto que son útiles cuando los gradientes están cerca de cero. En este trabajo seleccionamos la versión de ResNet50 que cuenta con más de 23 millones de parámetros. Las activaciones escogidas para codificar las características obtenidas por aprendizaje profundo fueron las de la capa *MaxPooling2D* que se encuentra después de la primera capa convolucional, y cuenta con un total de 64 filtros ³⁵.

- **Inception-ResNet-V2:** Es una arquitectura ResNet más sofisticada que fusiona bloques residuales con módulos inception que operan a través de activaciones convolucionales guardando la información estructural, y aprendiendo correlaciones no lineales. También, las convoluciones 1×1 , de los módulos inception, permiten disminuir el número de parámetros de entrenamiento de la red, sin embargo presentan más de 54 millones de parámetros, un número mayor al que contiene la arquitectura ResNet50. La capa seleccionada consta de 64 filtros ³⁶.
- **Densely Connected Convolutional Networks (DenseNet):** Su principal característica es el uso de concatenaciones profundas donde cada capa recibe como entrada la salida (mapas de activación) de todas las capas precedentes. Esta arquitectura aborda diversos problemas como el desvanecimiento del gradiente, la propagación se refuerza a la entrada de cada capa, ya que recibe

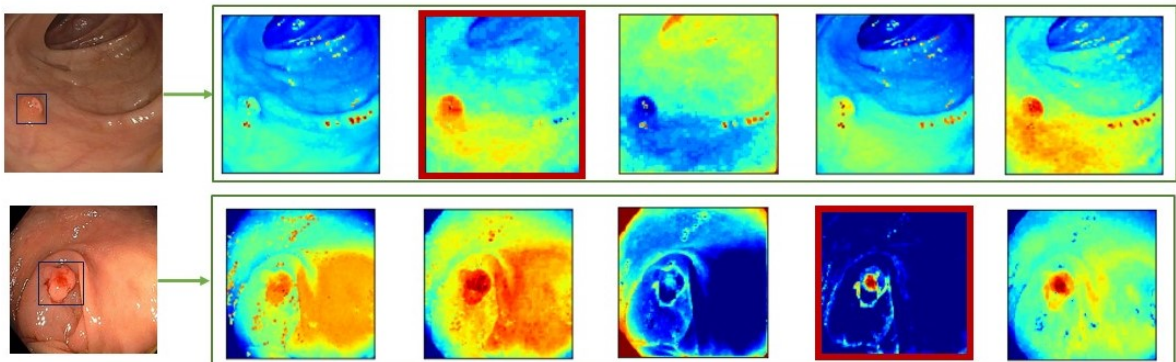
³⁵ HE, Kaiming, *et al.* "Deep residual learning for image recognition". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 770-778.

³⁶ SZEGEDY, Christian, *et al.* "Inception-v4, inception-resnet and the impact of residual connections on learning". En: *Thirty-first AAAI conference on artificial intelligence*. 2017.

información de todas las capas anteriores, y se reduce el número de parámetros de la red. En este trabajo se implementó la DenseNet201, que cuenta con más de 18 millones de parámetros y la capa después de la primera concatenación (con 64 filtros) fue seleccionada para codificar las características obtenidas por aprendizaje profundo durante la colonoscopia ³⁷.

En la figura 5 se evidencian los mapas de características obtenidos de la arquitectura convolucional ResNet50. Se puede apreciar que las diversas activaciones modelan la forma no paramétrica del pólipo, logrando diferenciar esta región de las estructuras similares presentes en el tracto intestinal como los pliegues e independientemente de la apariencia o ubicación del pólipo, este es resaltado, tal como se evidencia en las imágenes enmarcadas en rojo.

Figura 5. Mapas de características extraídas de la capa *MaxPooling2D* que se encuentra después de la primera capa convolucional de la arquitectura ResNet50. Cada fila representa los mapas de características obtenidos a partir de una protuberancia diferente. Es destacable que algunas activaciones centran su atención en el pólipo (cuadro rojo).



³⁷ HUANG, Gao , et al. *Densely Connected Convolutional Networks*. 2016. arXiv: 1608 . 06993 [cs.CV].

4.2. CONJUNTO DE DATOS

La evaluación del método propuesto se llevó a cabo sobre colonoscopias públicas presentadas en *Endoscopic Vision Challenge* (EndoVis). Se tuvieron en cuenta dos conjuntos de datos diferentes que se describen a continuación:

- *ASU-Mayo Clinic Colonoscopy Video Database* ²⁶ que cuenta con 20 videos de entrenamiento y 18 de prueba provenientes de procedimientos de colonoscopia con inspecciones cuidadosas o detalladas del colon (largas) y otras más generalizadas (cortas), capturados en el Departamento de Gastroenterología de la Clínica Mayo, Arizona. Este conjunto de datos reporta un único pólipo a lo largo de la secuencia, contemplando fuertes variaciones de luz y movimientos bruscos de la cámara, comunes en las colonoscopias. El conjunto de entrenamiento tiene un total de 3.799 fotogramas con pólipo, mientras que los de prueba cuentan con un total de 4.313 fotogramas con pólipo. Las diferentes protuberancias presentes en los fotogramas de colonoscopia son delineados por un experto en gastroenterología y están disponibles para la fase de entrenamiento y prueba. Cabe destacar que para la evaluación del método se usaron tanto las secuencias de entrenamiento como las de prueba.
- *CVC-Video Clinic DB database* ^{38, 39} tiene un total de 18 videos de colonoscopia con inspecciones tanto cortas como largas, con más de 12.613 fotogramas con pólipo. Estas colonoscopias fueron capturadas en la exploración rutinaria del Hospital Clínico de Barcelona, España. En cada una de las secuencias

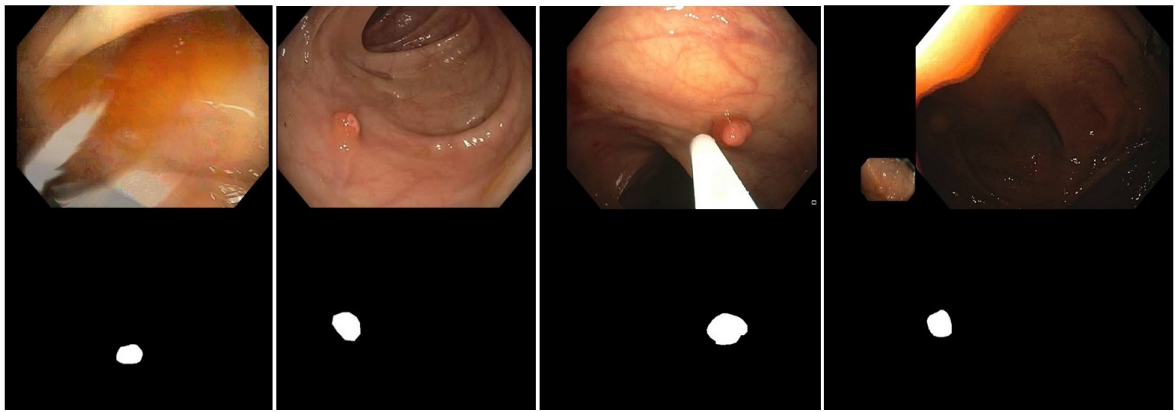
³⁸ BERNAL, Jorge, *et al.* "Polyp detection benchmark in colonoscopy videos using gcreator: A novel fully configurable tool for easy and fast annotation of image databases". En: 2018.

³⁹ ANGERMANN, Quentin, *et al.* "Towards Real-Time Polyp Detection in Colonoscopy Videos: Adapting Still Frame-Based Methodologies for Video Sequences Analysis". En: *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures*. Springer, 2017, págs. 29-41.

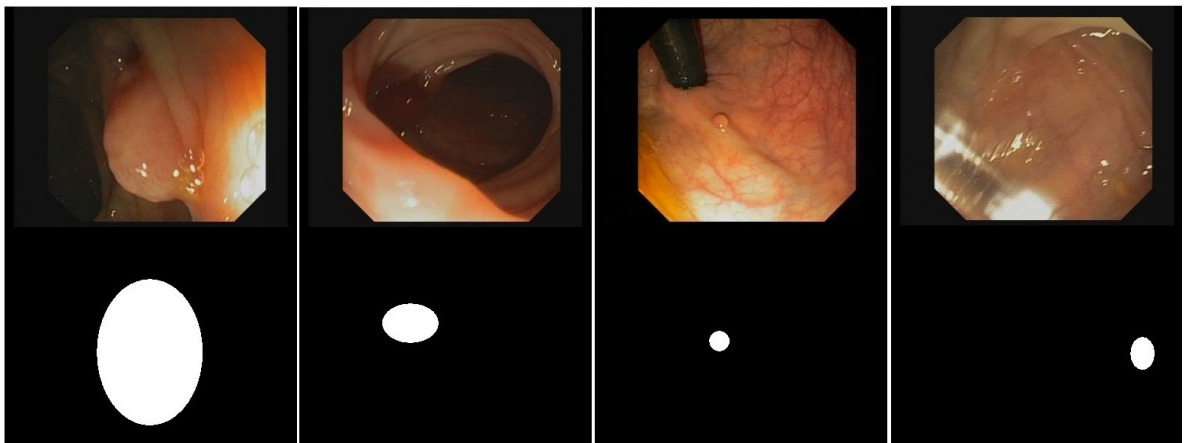
se reporta un único pólipo, el cual es identificado, ubicado y segmentado por expertos clínicos. Dado que el objetivo principal de esta labor es evaluar la capacidad del enfoque propuesto para recuperar los mapas de atención, sólo se han tenido en cuenta los videos del conjunto de entrenamiento que tienen la respectiva anotación y ubicación del pólipo.

Como ejemplo de los diferentes pólipos y marcaciones de cada conjunto de datos, se presenta la figura 6.

Figura 6. Conjunto de datos para realizar la evaluación del método. (a) *ASU-Mayo Clinic Colonoscopy Video Database*. (b) *CVC-Video Clinic DB database*.



(a)



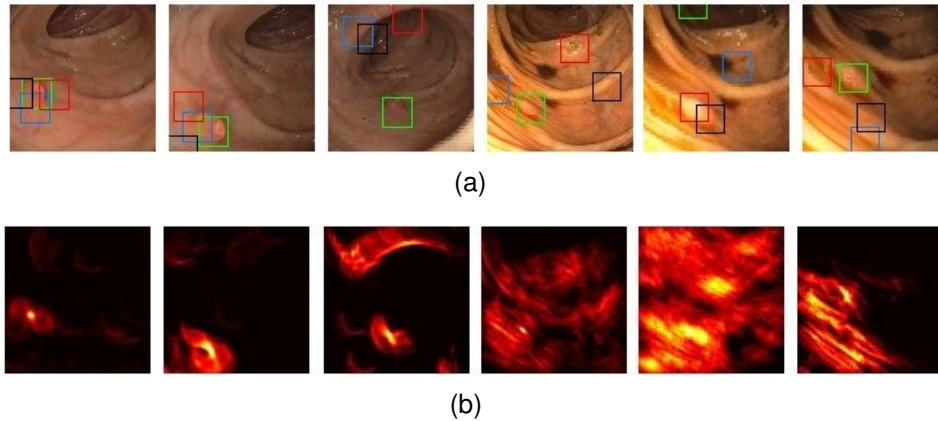
(b)

5. EVALUACIÓN Y RESULTADOS

Este trabajo está dedicado a generar mapas de atención de pólipos que permiten visualizar aquellas zonas del tracto intestinal con mayor probabilidad de tener estos biomarcadores durante un procedimiento rutinario de colonoscopia. Dichos mapas de atención se representan mediante la matriz Γ_t , siendo esta, el resultado del proceso de votación acumulativo, el cual proyecta la tabla representativa del pólipo (*R-Table*) sobre cada fotograma. Cada posición en Γ_t representa la probabilidad de ser el centro de una protuberancia presente en el tracto intestinal. La figura 7 describe los fotogramas típicos de colonoscopia (primera fila) con su respectivo mapa de atención Γ_t (segunda fila). Se puede visualizar en esta figura que las imágenes de colonoscopia tienen fuertes cambios en cuanto a la apariencia e iluminación. Además, los pólipos reportan una alta variabilidad ya sea en su forma o apariencia, teniendo una gran similitud con los pliegues intestinales presentes en las regiones circundantes.

En los mapas de atención, las regiones amarillas representan la ubicación con mayor probabilidad de coincidencia de los pólipos, a pesar de los desafíos ya mencionados, estos mapas se centran principalmente en las masas anormales. Como se ha observado, para los fotogramas que no presentan fuertes movimientos de la cámara (primeras tres columnas), el mapa de atención calculado enfoca sus votos en la región donde se encuentra el pólipo, con algunas probabilidades distribuidas a los pliegues que comparten características de apariencia y forma. En las otras tres columnas, los fuertes cambios de iluminación dificultan la tarea de detección de los pólipos, y los mapas de atención resultan afectados en la distribución de la energía, siendo esta, la concentración de los votos en un determinado mapa Γ_t .

Figura 7. Detección del pólipo en diferentes fotogramas y su respectiva matriz acumulativa Γ_t . (a) Detección proyectada a cada fotograma extrayendo cuatro regiones máximas de cada Γ_t . (b) Mapa de atención Γ_t para cada fotograma.

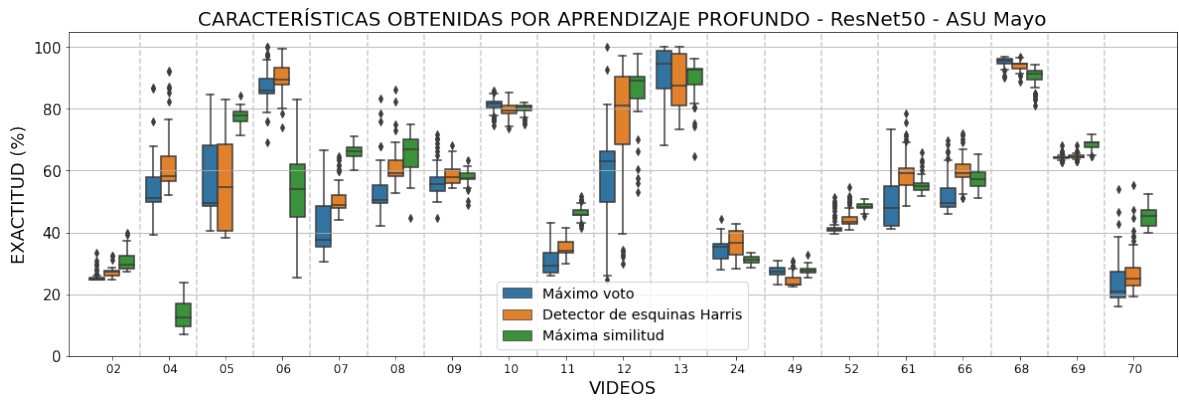


Los mapas de atención Γ_t fueron utilizados para buscar regiones que maximicen la probabilidad y así generar alarmas con respecto a la ubicación del pólipo, apoyando la tarea de detección durante la colonoscopia. Como se observa en la figura 7, aquellas regiones máximas extraídas de cada mapa de atención son proyectadas a los fotogramas (cajas de color). En este caso, se seleccionaron cuatro regiones, en orden descendente de importancia se encuentra la caja verde, roja, azul y negra. Es destacable que incluso en aquellos mapas de atención que contienen ruido, las regiones seleccionadas coinciden correctamente con protuberancias. En todos los casos, el pólipo se encuentra encerrado por alguna de las regiones detectadas. Se diseñó un primer experimento para evaluar el rendimiento del método propuesto al momento de recuperar la ubicación más probable del pólipo en cada fotograma de las secuencias. Para ello, se generaron mapas de atención Γ_t sobre las colonoscopias, y luego, la ubicación más probable se proyectó en cada fotograma siguiendo las estrategias de máximo voto, similitud, y la ubicación del pólipo a partir del detector de esquinas de Harris (sección 3.3). Se implementó una medida de superposición para encontrar la correspondencia con respecto a las anotaciones de cada secuencia realizadas por un gastroenterólogo experto. En cuanto a la valida-

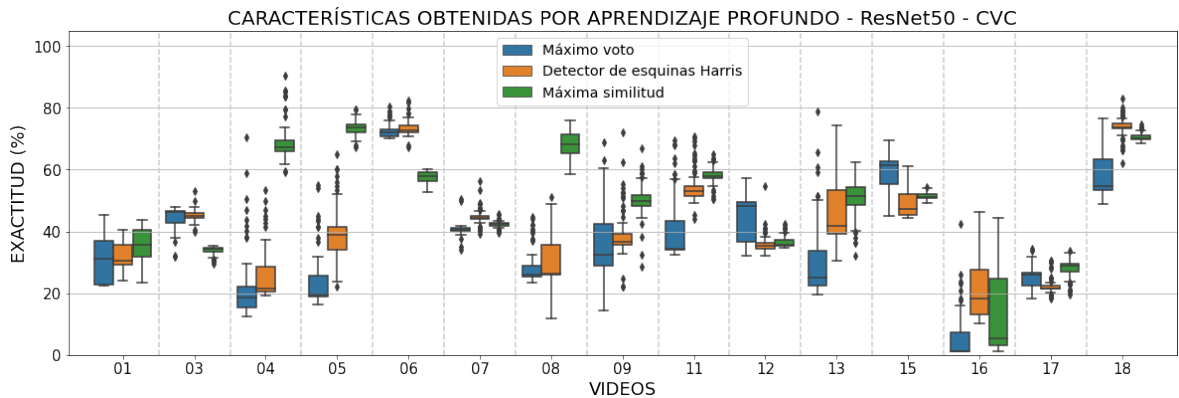
ción, si el punto máximo proyectado cae en cualquier lugar dentro de la anotación, entonces se considera como una verdadera detección (VP). La exactitud entonces se define como: $\#VP/\#muestras$. La figura 8 resume los resultados obtenidos para cada secuencia considerada en el conjunto de datos *ASU-Mayo* y *CVC* con la red convolucional ResNet50, ya que esta arquitectura obtuvo la mejor representación del pólipo en cada una de sus activaciones. En estas gráficas se ilustran los resultados alcanzados para cada secuencia con respecto a las estrategias de extracción de máximos implementadas. Como se ha observado, la estimación del pólipo es una tarea difícil, capturada en el mapa de atención Γ_t , pero no necesariamente como el valor más probable. De hecho, la regla de localización de Harris-ResNet50 obtuvo los mejores resultados con un promedio de incidencia de 58 % y 42 % para el conjunto de datos *ASU-Mayo* y *CVC*, respectivamente. Asimismo, cabe destacar un notable rendimiento en las secuencias del conjunto de datos *ASU-Mayo*, obteniendo porcentajes de hasta el 90 % de puntaje de detección en algunos videos (*06*, *12*, *13*). Esta exitosa detección se asocia a una representación profunda y robusta que logra una descripción completa de los pólipos. Además, estas secuencias comparten que los pólipos son salientes en el tracto intestinal, mientras que otros videos (como *02*, *24*, *49* de *ASU-Mayo*) reportan pólipos inapreciables o similares a las estructuras presentes en el tracto intestinal, siendo difíciles de localizar, incluso para un gastroenterólogo experto. El mismo experimento se llevó a cabo a partir de otras arquitecturas convolucionales, como Inception-ResNetV2 y DenseNet201, donde la primera logra resultados promedios de 49 % para el conjunto de datos *ASU-Mayo* y 40 % para *CVC*, por su parte la arquitectura DenseNet201 alcanza resultados de 56 % para el conjunto de datos *ASU-Mayo* y 42 % para *CVC*, con la regla de detección de esquinas de Harris. Como se observa, estos resultados son menores que los obtenidos por ResNet50. En ese caso, las activaciones residuales lograron una mejor descripción del fotograma que las otras arquitecturas, especialmente cuando

los objetos comparten características muy parecidas como lo es el fondo de cada fotograma.

Figura 8. Resultados evaluando los operadores locales de búsqueda de máximos para seleccionar las regiones de pólipos más probables con el método de extracción de características obtenidas por aprendizaje profundo. Se presentan con la arquitectura ResNet50. (a) *ASU-Mayo*. (b) *CVC*.



(a)

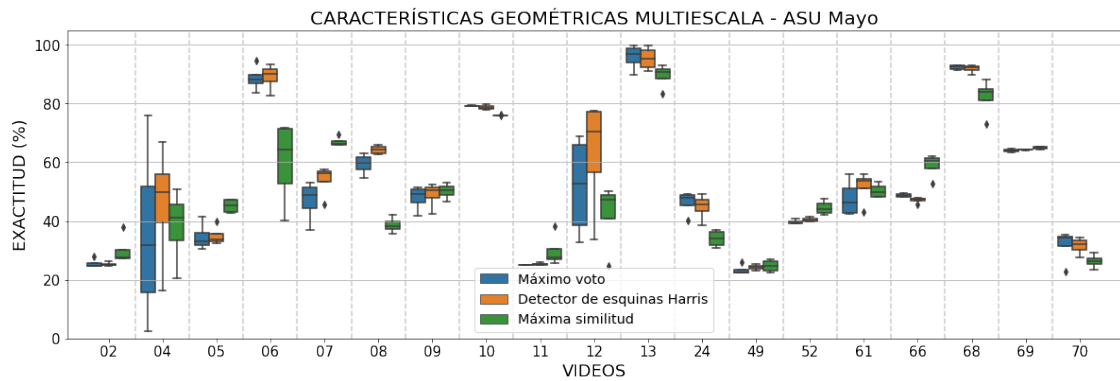


(b)

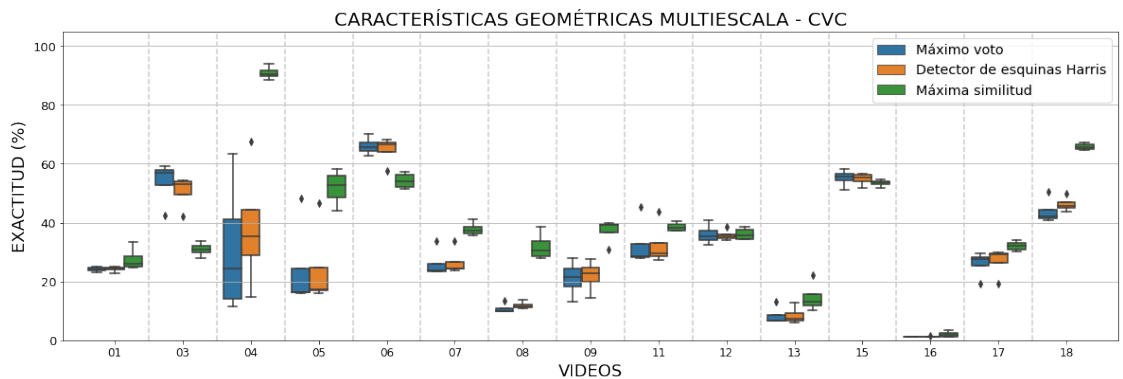
La figura 9 reporta los resultados obtenidos para las características geométricas multi-escala. Como era de esperar, la escasa representación centrada en los puntos salientes alrededor de los pólipos resulta menos representativo para detectar y localizar los mismos durante la colonoscopia. De hecho, el uso de esta representación obtuvo 5% y 13% menos que el método de características obtenidas por apren-

dizaje profundo para los conjunto de datos *ASU-Mayo* y *CVC*, respectivamente. El detector de Harris fue nuevamente la mejor estrategia para extraer las regiones más precisas de cada mapa de votación, reportando aproximadamente 53% y 29% de puntaje de detección para los conjuntos de datos *ASU-Mayo* y *CVC*. En esta estrategia, cada diagrama de caja presenta una pequeña variación atribuida a la representación de baja dimensión, ya que cada *R-Table* aprende sólo tres características, mientras que con las características obtenidas por aprendizaje profundo se obtienen 64 activaciones, debido a esto se puede inferir que el tener una mayor cantidad de datos puede conducir a una mayor variación en las detecciones.

Figura 9. Resultados evaluando los operadores locales de búsqueda de máximos para seleccionar las regiones de pólipos más probables con el método de extracción de características geométricas multi-escala. (a) *ASU-Mayo*. (b) *CVC*.



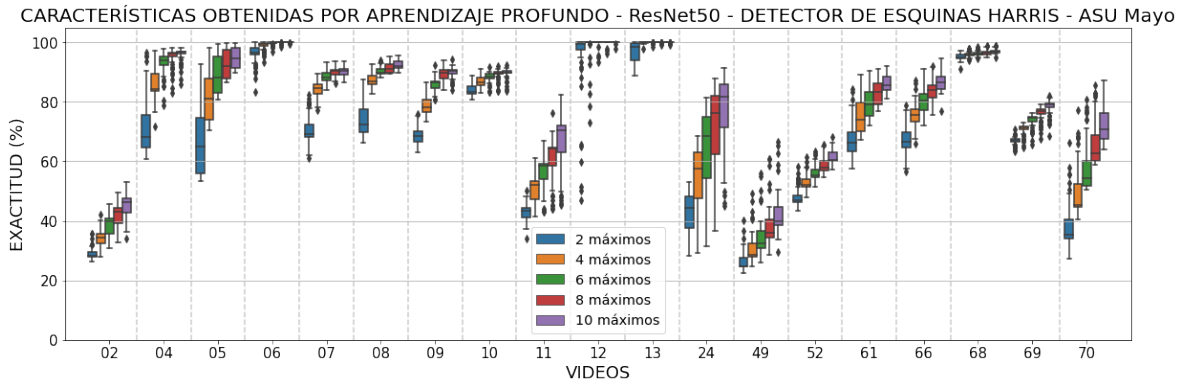
(a)



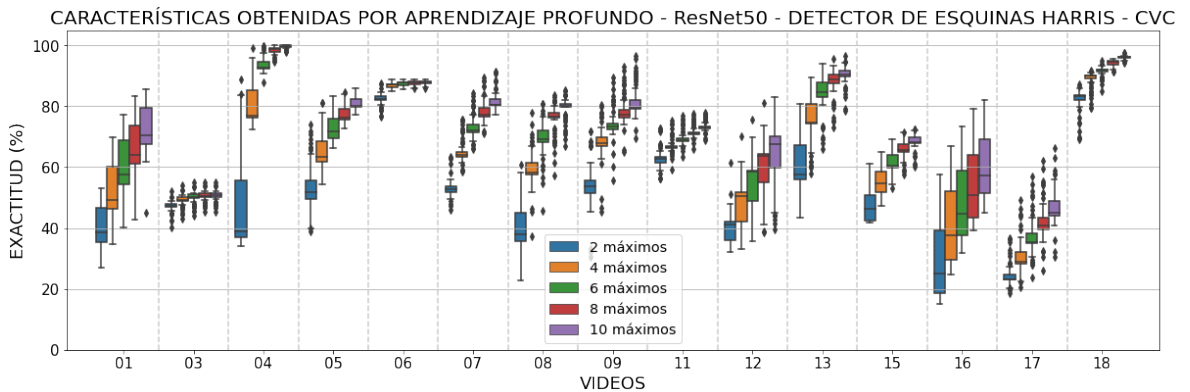
(b)

Se implementó un segundo experimento para seleccionar un número reducido de puntos relevantes en el mapa Γ_t y luego proyectarlos para obtener regiones (Θ) en los fotogramas de colonoscopia. Este experimento tiene por objetivo emular el enfoque propuesto como sistema de alerta de pólipos que apoya la tarea de detección y localización durante el procedimiento clínico. La configuración de los máximos elegidos en cada mapa de atención fue $\Theta = \{2, 4, 6, 8, 10\}$. La figura 10 resume los resultados alcanzados usando características obtenidas por aprendizaje profundo en el conjunto de datos *ASU-Mayo* y *CVC*. Como era de esperar, la proyección de un gran número de regiones permite encontrar correctamente el pólipo en las diferentes secuencias, siendo la tarea más relevante durante el procedimiento clínico (precisión media superior al 70%). De hecho, para algunos videos (como *05*, *06*, *12*, *13* de *ASU-Mayo*) se logra una perfecta detección del pólipo en alguna de las regiones seleccionadas del mapa de atención Γ_t . Cabe señalar también, que otras regiones se detectan erróneamente, pero en este trabajo, los mapas Γ_t están diseñados para centrar la atención del gastroenterólogo experto, y así describir o encontrar algún patrón anormal en los pliegues a lo largo del tracto intestinal. En promedio, un total de cuatro regiones parecen suficientes para delimitar un pólipo a partir de los mapas de atención, con algunas diferencias para los videos que reportan fuertes movimientos de la cámara y presentan diversos artefactos que se encuentran durante la colonoscopia. Por ejemplo, los videos denominados *03* y *12* en *CVC* reportan grandes desafíos para localizar la región del pólipo y pueden requerir un gran número de regiones para encerrar adecuadamente dicha región. En otras secuencias, como la *24* en *ASU-Mayo* y *16* en *CVC*, reportan una mayor variación en la detección de los pólipos a lo largo de la secuencia de video, con un número diferente de regiones. Este fenómeno se asocia a los movimientos bruscos de la cámara y a oclusiones del pólipo que afectan la detección.

Figura 10. Resultados evaluando diferentes máximos Θ en cada mapa de atención Γ_t con la estrategia de extracción de características obtenidas por aprendizaje profundo y la arquitectura convolucional ResNet50. (a) *ASU-Mayo*. (b) *CVC*.



(a)

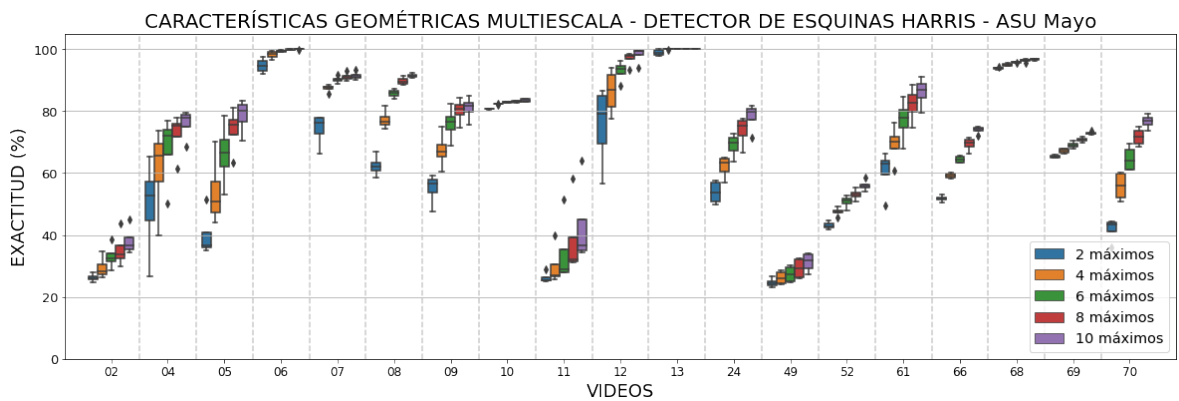


(b)

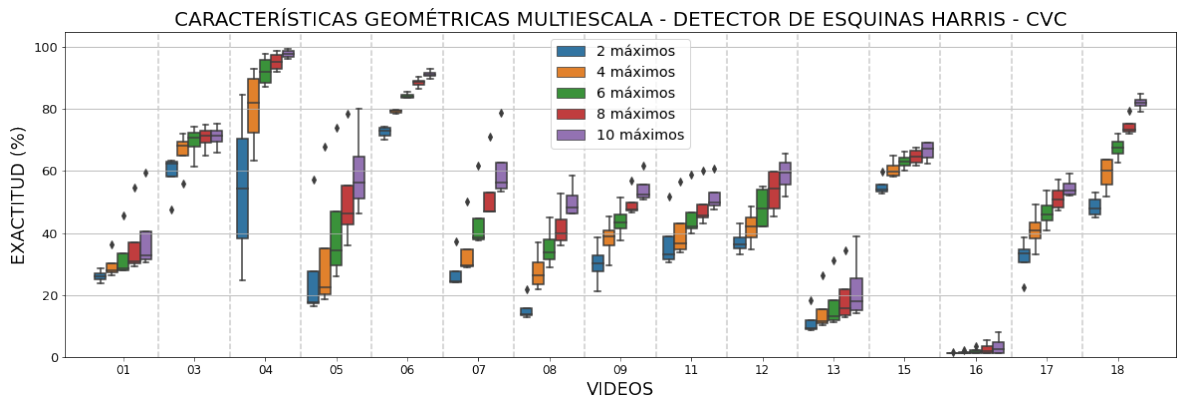
En la misma línea, se obtuvo la proyección de varios puntos salientes en los mapas de atención Γ_t a los fotogramas de las secuencias de colonoscopia, pero utilizando las características geométricas multi-escala para representar los pólipos. En la figura 11 se ilustran los resultados obtenidos para ambos conjuntos de datos: *ASU-Mayo* y *CVC*. Como era de esperar, la estimación de un mayor número de regiones aumenta la probabilidad de encerrar estos biomarcadores, lo que muestra la capacidad robusta de los mapas de atención para realzar puntos con características similares a las del pólipo aprendido. No obstante, esta representación tiene una precisión

menor que la lograda con las características obtenidas por aprendizaje profundo, y para gran parte de los videos pueden ser necesarias un mayor número de regiones para detectar adecuadamente estas masas anormales. Las principales limitaciones se señalan en el conjunto de datos *CVC*, que tiene una precisión media de 55% utilizando diez regiones sobre las colonoscopias. En particular, la representación de estas características es insuficiente en la secuencia 16 de *CVC* para tratar los movimientos bruscos y las oclusiones de los pólipos.

Figura 11. Resultados evaluando diferentes máximos Θ en cada mapa de atención Γ_t con la estrategia de extracción de características geométricas multi-escala. (a) *ASU-Mayo*. (b) *CVC*.



(a)

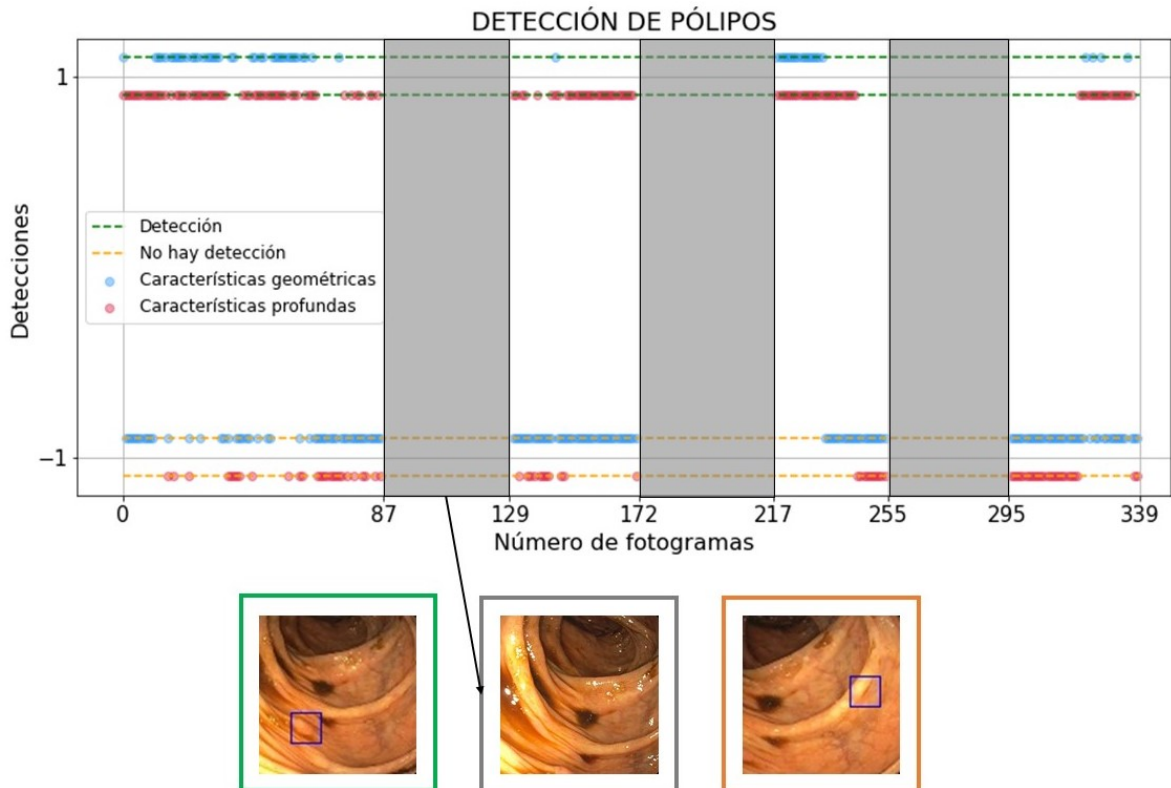


(b)

La capacidad del método propuesto para la detección de los pólipos puede cambiar a lo largo de la secuencia de video debido a la oclusión, el cambio de iluminación durante la colonoscopia, los diversos artefactos (pliegues intestinales, reflejos especulares), entre otras muchas variables. Por ejemplo, en la figura 12 se ilustra una secuencia de colonoscopia típica que tiene un pólipo identificado, el cual podría estar totalmente ocluido o desaparecer de la escena en algunos intervalos de fotogramas (zona gris). La línea horizontal verde y la naranja representan una detección positiva o verdadera y una detección errónea, respectivamente. En este ejemplo, la representación de la *R-Table* con la estrategia de las características obtenidas por aprendizaje profundo (puntos rojos) muestra un rendimiento más estable incluso después de las zonas que la protuberancia está totalmente ocluida (regiones grises), recuperando rápidamente una detección coherente del pólipo después de estos intervalos. En cambio, la representación de características geométricas multi-escala (puntos azules) reporta más dificultades para recuperar la detección del pólipo, presentando grandes variaciones a lo largo de la secuencia, e incluso, en algunos intervalos reporta un mayor número de detecciones erróneas que verdaderas. Por otra parte, la detección de pólipos es llevada a cabo en cada fotograma, bajo un análisis de fotogramas independientes (sin dependencia temporal), que evita clasificaciones erróneas producidas por observaciones ruidosas de la cámara. Del mismo modo, pequeños saltos y movimientos acelerados de la cámara producen distorsiones temporales, por lo cual se justifican observaciones independientes.

A continuación se llevó a cabo una evaluación cuantitativa para medir la capacidad del enfoque propuesto a lo largo de las secuencias de video. Para esto, se definió una detección de pólipo positiva (VP) cuando el método delimitó adecuadamente una región de pólipo teniendo en cuenta cuatro regiones en cada mapa de votación Γ_t . A partir de este valor (VP) se estableció una métrica de precisión definida como: $P(t) = \#VP/\#\text{pólipos}(t)$. Esta evaluación temporal, $P(t)$ cuenta la tasa de pólipos

Figura 12. Detección de pólipos en las secuencias de colonoscopia. Como se ha observado, los pólipos cambian a lo largo de la secuencia, y en algunos casos, pueden pasar inadvertidos, incluso para los gastroenterólogos expertos. La línea horizontal verde y la naranja representan una detección positiva y una detección errónea, respectivamente. Las zonas grises definen aquellos intervalos en los que el pólipo está totalmente ocluido desapareciendo de la escena.



verdaderos detectados con respecto al incremento de muestras de pólipos en tal intervalo del video. La figura 13 ilustra este rendimiento temporal en promedio obtenido para los conjuntos de datos *ASU-Mayo* (triángulo rojo) y *CVC* (triángulo verde), usando las características obtenidas por aprendizaje profundo, mientras que la figura 14 muestra el resultado obtenido con las características geométricas multi-escala para los conjuntos de datos *ASU-Mayo* (triángulo rojo) y *CVC* (triángulo verde), en la que el eje de las abscisas representa el porcentaje de fotogramas evaluados y el eje de las ordenadas la puntuación obtenida. Como era de esperar, las caracterís-

Las características obtenidas por aprendizaje profundo tienen un rendimiento considerablemente mejor, presentando una mejor media en la precisión para ambos conjuntos de datos. Cabe señalar que en este caso la precisión se mantiene estable (entre el 70 % y 90 %) a lo largo de las secuencias de colonoscopia del conjunto de datos *ASU-Mayo* mientras que con *CVC* la precisión se mantiene entre un 60-80 %, con pequeñas diferencias debido a las regiones ocluidas (figura 13). En contraste, el enfoque basado en las características geométricas multi-escala tiene una precisión limitada debido a la escasa representación del pólipo siendo vulnerable a la oclusión y a los fuertes movimientos de cámara, principalmente para el conjunto de datos *CVC* donde la precisión se mantiene en un rango de 25-45 % en promedio mientras que el conjunto de datos *ASU-Mayo* se encuentra en un rango de 60-80 % (figura 14).

Figura 13. Resultados basados en fotogramas para las características obtenidas por aprendizaje profundo. En los gráficos se observa la precisión, calculada recursivamente a lo largo de cada secuencia de los conjuntos de datos *ASU-Mayo* y *CVC*.

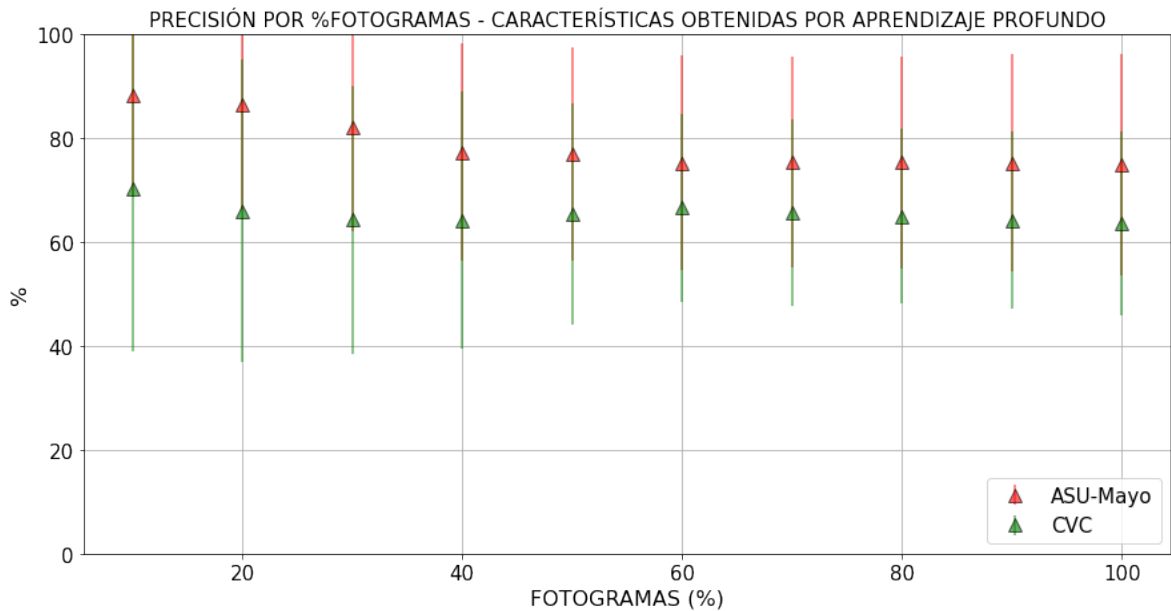
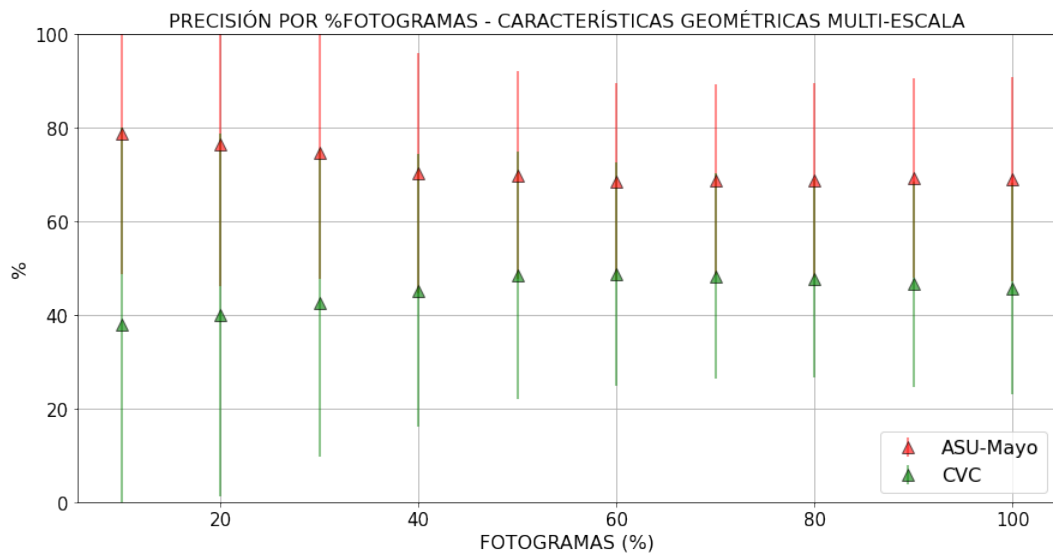


Figura 14. Resultados basados en fotografías para las características geométricas multi-escala. En los gráficos se observa la precisión, calculada recursivamente a lo largo de cada secuencia de los conjunto de datos *ASU-Mayo* y *CVC*.



6. CONCLUSIONES Y PERSPECTIVAS

Este trabajo presentó un enfoque computacional novedoso que genera mapas de atención visual en cada fotograma durante una colonoscopia, el cual realza las regiones sospechosas relacionadas con las masas propias del cáncer de colon. Estos mapas se construyen a partir de la transformada densa de Hough que aprende características multi-dimensionales para cada píxel, relacionados con una plantilla, que delimita la muestra de pólipo. En el enfoque propuesto, se implementaron dos métodos: 1) las características geométricas multi-escala y 2) las las características obtenidas por aprendizaje profundo, para modelar el pólipo en las tablas representativas (*R-Tables*). Entonces, para cada fotograma, cada píxel es mapeado a una *R-Table* donde se almacena su distancia con respecto a la posición estimada del pólipo. Los mapas de atención visual obtienen un gran número de votos sobre la ubicación real del pólipo, identificándolo incluso cuando hay oclusión en fotogramas intermedios. Con este método se logra un rendimiento robusto en diferentes secuencias de colonoscopia, requiriendo sólo el primer fotograma para aprender varias características que representan la variación del pólipo de manera adecuada. La herramienta propuesta puede apoyar la detección y caracterización de masas anormales durante la colonoscopia con la capacidad para funcionar en línea durante el procedimiento clínico a partir de mapas de atención, que podrían visualizarse en paralelo con los fotogramas de la secuencia. También es posible computar la ubicación máxima y proyectar las regiones extraídas a la colonoscopia original para realzar las diversas regiones anormales. A pesar de que en muchas de estas regiones no existen pólipos, el método propuesto para enfocar la atención del especialista durante el procedimiento clínico mediante los mapas de atención podría mitigar la falta de detección de estos biomarcadores, la cual se estima entre un 4-12% ^{8, 9} e incluso puede llegar hasta un 25-26% según últimos estudios clínicos realizados ^{10, 11, 12}.

En trabajos similares se ha explorado la detección e identificación de pólipos durante las colonoscopias. En cuanto a la detección, varios trabajos han modelado independientemente algunas primitivas tales como el color, los bordes y la textura ²⁰. Sin embargo, estos enfoques se limitan a representar toda la complejidad y variabilidad de los pólipos en el procedimiento de colonoscopia. En contraste, el enfoque propuesto obtuvo una representación del pólipo a partir de los métodos de extracción de características anteriormente mencionados. De este proceso se puede concluir que el implementar características obtenidas por aprendizaje profundo permite un modelado más robusto por medio de la codificación de un banco de filtros en la *R-Table*, aprendidos de imágenes provenientes de un dominio natural (ImageNet) y teniendo como base arquitecturas pre-entrenadas. Una de las principales ventajas de estos filtros es la separación de la banda a partir de funciones no lineales que pueden potencialmente separar mejor los patrones primitivos de los pólipos. Este método logró un modelado de los pólipos más estable y con mejor desempeño a lo largo de las diferentes secuencias. Otros enfoques también han implementado la transformada de Hough, pero utilizando una perspectiva escasa y clásica, asumiendo el pólipo como una forma paramétrica ²². Esta suposición resulta poco realista, especialmente porque la observación del pólipo es altamente no paramétrica, debido a los pliegues en el tracto intestinal y al movimiento de la cámara. Recientemente, los enfoques de aprendizaje profundo se han utilizado en tareas de detección de pólipos. Por ejemplo, Zheng *et al.* (2019) ²⁸ propuso una arquitectura similar a la *U-Net* que junto con el flujo óptico muestra resultados prometedores en un conjunto de datos académicos. Este trabajo aprende una robusta representación del pólipo, pero requiere un alto número de secuencias de colonoscopia, para hacer frente a las nuevas observaciones de pólipos comunes en la rutina clínica. En nuestro caso, aprovechamos las representaciones profundas previamente entrenadas, que resultan robustas para representar la variabilidad general de la imagen. Cabe resaltar

que en el enfoque propuesto, el reto reside en la detección de regiones máximas, ya que para un conjunto de regiones, el método logra una perfecta puntuación en la detección del pólipo sobre conjuntos de datos públicos. Además, los mapas de atención pueden centrarse en otras regiones con patrones de pólipos similares, que podrían ser de interés para el gastroenterólogo experto.

Durante la evaluación se informó de varias limitaciones del método propuesto para abordar la compleja estructura del tracto intestinal, tales como los movimientos bruscos de la cámara y las diferentes irregularidades visuales que corresponden a los pliegues intestinales o artefactos como los reflejos especulares. Esas alteraciones de la observación durante la colonoscopia pueden contribuir a realzar las regiones de falsos positivos en los mapas de atención. Como trabajo futuro, se propone el análisis y aprendizaje de las características específicas de los pólipos para lograr mapas de atención más centrados en las protuberancias de interés. La detección totalmente automática de los pólipos se evaluará mediante el aprendizaje de un atlas de pólipos que podría ser adecuado para cada una de las secuencias de colonoscopia de prueba. También se prevé desarrollar un estudio en una rutina clínica para medir la contribución de esta herramienta.

BIBLIOGRAFÍA

ALAGAPPAN, Muthuraman, *et al.* “Artificial intelligence in gastrointestinal endoscopy: The future is almost here”. En: *World journal of gastrointestinal endoscopy* 10.10 (2018), pág. 239 (vid. pág. 25).

AMELING, Stefan, *et al.* “Texture-Based Polyp Detection in Colonoscopy”. En: *Bildverarbeitung für die Medizin 2009*. Ed. por Hans-Peter MEINZER, *et al.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, págs. 346-350 (vid. pág. 14).

ANGERMANN, Quentin; HISTACE, Aymeric y ROMAIN, Olivier. “Active learning for real time detection of polyps in videocolonoscopy”. En: *Procedia Computer Science* 90 (2016), págs. 182-187 (vid. págs. 12, 48).

ANGERMANN, Quentin, *et al.* “Towards Real-Time Polyp Detection in Colonoscopy Videos: Adapting Still Frame-Based Methodologies for Video Sequences Analysis”. En: *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures*. Springer, 2017, págs. 29-41 (vid. pág. 34).

BASAVE, Horacio Noé López, *et al.* “Cáncer de colon y recto”. En: *Manual de Oncología, 6e*. Ed. por Ángel HERRERA-GÓMEZ; Silvio A. Ñamendys SILVA y Abelardo MENESES-GARCÍA. New York, NY: McGraw-Hill Education, 2018 (vid. pág. 10).

BERNAL, J., *et al.* “Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge”. En: *IEEE Transactions on Medical Imaging* 36.6 (2017), págs. 1231-1249 (vid. pág. 12).

BERNAL, Jorge; SÁNCHEZ, Javier y VILARINO, Fernando. "Towards automatic polyp detection with a polyp appearance model". En: *Pattern Recognition* 45.9 (2012), págs. 3166-3182 (vid. pág. 13).

BERNAL, Jorge, *et al.* "Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge". En: *IEEE transactions on medical imaging* 36.6 (2017), págs. 1231-1249 (vid. págs. 13, 49).

BERNAL, Jorge, *et al.* "Polyp detection benchmark in colonoscopy videos using gt-creator: A novel fully configurable tool for easy and fast annotation of image databases". En: 2018 (vid. pág. 34).

BERTALMIO, Marcelo; BERTOZZI, Andrea L y SAPIRO, Guillermo. "Navier-stokes, fluid dynamics, and image and video inpainting". En: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. IEEE. 2001, págs. I-I (vid. pág. 26).

BRESSLER, Brian, *et al.* "Rates of New or Missed Colorectal Cancers After Colonoscopy and Their Risk Factors: A Population-Based Analysis". En: *Gastroenterology* 132.1 (2007), págs. 96 -102. DOI: <https://doi.org/10.1053/j.gastro.2006.10.027> (vid. págs. 12, 48).

CLERCQ, Chantal MC le, *et al.* "Postcolonoscopy colorectal cancers are preventable: a population-based study". En: *Gut* 63.6 (2014), págs. 957-963 (vid. pág. 12).

DAVILA, Raquel E.; RAJAN, Elizabeth y BARON, Todd H. "ASGE guideline: colorectal cancer screening and surveillance". En: *Gastrointestinal Endoscopy* 63.4 (2006), págs. 546 -557. DOI: <https://doi.org/10.1016/j.gie.2006.02.002> (vid. pág. 11).

FERLAY J Colombet M, Soerjomataram I Mathers C Parkin DM Piñeros M Znaor A Bray F. “Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods doi: 10.1002/ijc.31937. [Epub ahead of print] PubMed PMID: 30350310”. En: *Int J Cancer*. (2018 Oct 23) (vid. pág. 10).

GANZ, Melanie; YANG, Xiaoyun y SLABAUGH, Greg. “Automatic segmentation of polyps in colonoscopic narrow-band imaging data”. En: *IEEE Transactions on Biomedical Engineering* 59.8 (2012), págs. 2144-2151 (vid. pág. 13).

HE, Kaiming , *et al.* “Deep residual learning for image recognition”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 770-778 (vid. pág. 32).

HUANG, Gao, *et al.* *Densely Connected Convolutional Networks*. 2016. arXiv: 1608.06993 [cs.CV] (vid. pág. 33).

IAKOVIDIS, Dimitrios K, *et al.* “A comparative study of texture features for the discrimination of gastric polyps in endoscopic video”. En: *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*. IEEE. 2005, págs. 575-580 (vid. pág. 14).

IWAHORI, Yuji, *et al.* “Automatic detection of polyp using Hessian Filter and HOG features”. En: *Procedia computer science* 60 (2015), págs. 730-739 (vid. pág. 14).

JESSICA COHAN, MD y MADHULIKA G VARMA, MD. “Intestino grueso”. En: *Diagnóstico y tratamiento quirúrgicos, 14e*. Ed. por Gerard M. DOHERTY. New York, NY: McGraw-Hill Education, 2018 (vid. pág. 11).

KUPERIJ, N., *et al.* “Design of a user interface for intuitive colonoscope control”. En: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2011, págs. 937-942 (vid. pág. 11).

MANZANERA, Antoine. "Dense Hough transforms on gray level images using multi-scale derivatives". En: *SIXIEME WORKSHOP AMINA 2012 Applications Médicales de l'Informatique: Nouvelles Approches*. 2012 (vid. págs. 16, 20).

MO, X. , *et al.* "An Efficient Approach for Polyps Detection in Endoscopic Videos Based on Faster R-CNN". En: *2018 24th International Conference on Pattern Recognition (ICPR)*. 2018, págs. 3929-3934. DOI: 10.1109/ICPR.2018.8545174 (vid. pág. 15).

PEREZ, Eduardo. *Gastroenterologa*. McGraw Hill Mexico, 2012 (vid. págs. 10, 13).

POINTET, A-L y TAIEB, J. "Autoevaluaciones del artículo: Cáncer de colon". En: (2016) (vid. pág. 12).

RABENECK, Linda y PASZAT, Lawrence F. "Circumstances in which colonoscopy misses cancer". En: *Frontline gastroenterology* 1.1 (2010), págs. 52-58 (vid. pág. 12).

STEHLE, Thomas , *et al.* "Classification of Colon Polyps in NBI Endoscopy Using Vascularization Features". En: *Medical Imaging 2009: Computer-Aided Diagnosis*. Ed. por N. KARSSEMEIJER y M. L. GIGER. Vol. 7260. Orlando, USA: SPIE, 2009 (vid. pág. 13).

SZEGEDY, Christian, *et al.* "Inception-v4, inception-resnet and the impact of residual connections on learning". En: *Thirty-first AAAI conference on artificial intelligence*. 2017 (vid. pág. 32).

TAJBAKSH, N.; GURUDU, S. R. y LIANG, J. "Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information". En: *IEEE Transactions on Medical Imaging* 35.2 (2016), págs. 630-644 (vid. págs. 11, 48).

TAJBAKSH, Nima; GURUDU, Suryakanth R y LIANG, Jianming. “Automated polyp detection in colonoscopy videos using shape and context information”. En: *IEEE transactions on medical imaging* 35.2 (2016), págs. 630-644 (vid. págs. 14, 34).

— “Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks”. En: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2015, págs. 79-83 (vid. pág. 13).

TARIK, Ghannam, *et al.* “Polyps’s region of interest detection in colonoscopy images by using clustering segmentation and region growing”. En: *Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on*. IEEE. 2016, págs. 455-459 (vid. págs. 14, 49).

VAN DE SANDE, Koen; GEVERS, Theo y SNOEK, Cees. “Evaluating color descriptors for object and scene recognition”. En: *IEEE transactions on pattern analysis and machine intelligence* 32.9 (2010), págs. 1582-1596 (vid. pág. 23).

VAN RIJN, Jeroen C, *et al.* “Polyp miss rate determined by tandem colonoscopy: a systematic review”. En: *The American journal of gastroenterology* 101.2 (2006), pág. 343 (vid. págs. 12, 48).

WANG, Y., *et al.* “Part-Based Multiderivative Edge Cross-Sectional Profiles for Polyp Detection in Colonoscopy”. En: *IEEE Journal of Biomedical and Health Informatics* 18.4 (2014), págs. 1379-1389 (vid. págs. 11, 48).

ZHANG, Ruikai, *et al.* “Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker”. En: *Pattern Recognition* 83 (2018), págs. 209 -219. DOI: <https://doi.org/10.1016/j.patcog.2018.05.026> (vid. pág. 15).

ZHENG, H., *et al.* "Polyp Tracking in Video Colonoscopy Using Optical Flow With an On-The-Fly Trained CNN". En: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019, págs. 79-82. DOI: 10.1109/ISBI.2019.8759180 (vid. págs. 15, 49).

ANEXOS

Anexo A. Productos Académicos

- L. Ruiz, L. Guayacán, F. Martínez. “Automatic polyp detection from a regional appearance model and a robust dense Hough coding”. 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA).
Estado: publicado.
- L. Ruiz, L. Guayacán, F. Martínez. “Attention maps to detect and localize potential malign polyp masses during colonoscopy”. Computer Methods and Programs in Biomedicine.
Estado: en revisión.