

# LOCALIZACIÓN E IDENTIFICACIÓN AUTOMÁTICA DE PÓLIPOS MEDIANTE UNA RED NEURONAL CONVOLUCIONAL POR REGIONES

**Luna Perejon, Francisco\*;** Duran López, Lourdes; Amaya Rodríguez, Isabel R., Civit Masot, Javier; Civit Balcells, Antón; Vicente Díaz, Saturnino; Linares Barranco, Alejandro

Robotics and Technology of Computers.

Arquitectura y Tecnología de Computadores. Escuela Superior de Ingeniería Informática. Universidad de Sevilla.

\*E-mail: [fralunper@atc.us.es](mailto:fralunper@atc.us.es)

## RESUMEN

Este trabajo expone la metodología llevada a cabo para la aplicación de un modelo *Deep Learning* con el fin de detectar pólipos de forma automática, así como su posición en videos de colonoscopia. Se plantearon diferentes métodos y diversas técnicas que pudieran aplicarse sobre el conjunto de datos proporcionado por el *2018 Sub-challenge Gastrointestinal Image ANALysis*. Seleccionamos el método *Faster Regional Convolutional Neural Network* para abarcar el problema planteado. Para la extracción de características empleamos el modelo *ResNet50*. Aplicamos técnicas de *data augmentation* para incrementar el conjunto de datos empleado en el entrenamiento del modelo. También aplicamos *hard negative mining* para reforzar el aprendizaje del *background* o fondo, reducir el porcentaje de falsos positivos y mejorar el rendimiento.

## PALABRAS CLAVE

Colonoscopia, Detección de pólipos, R-CNN, Regiones de interés.

## ABSTRACT

This work exposes the methodology carried out for the application of a *Deep Learning* model in the context of automatic polyp detection and its location in colonoscopy videos. Different methods were proposed as well as the different techniques that can be applied on the given dataset provided by the *2018 Sub-challenge Gastrointestinal Image ANALysis*. We chose the *Faster Regional Convolutional Neural Network* method to solve this problem. We used *ResNet50* in the first part of this algorithm to extract the main image features. We applied *hard negative mining* and *data augmentation* techniques to increase the dataset used in the training of the model. We also used *hard negative mining* to get a better learning of background, reducing false negatives and improving the performance.

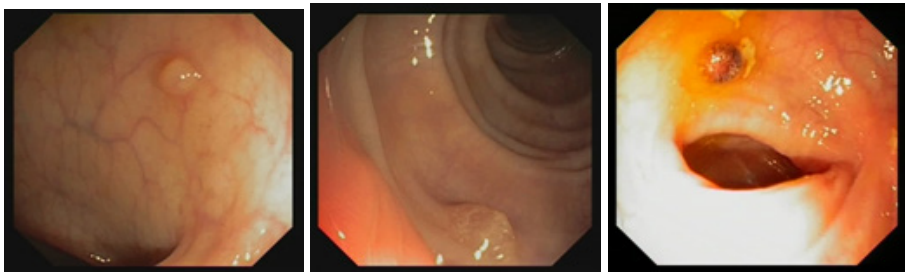
## KEYWORDS

Colonoscopy, Polyp detection, R-CNN, Regions of interest.

## INTRODUCCIÓN Y OBJETIVOS

Los algoritmos de *Deep Learning* (DL en adelante) en la última década han demostrado poseer un elevado rendimiento al aplicarlos a problemas de clasificación [1]. Esto ha causado que en la actualidad se erija como la tecnología emergente de mayor expectación [2]. Destacan principalmente en el análisis y procesado de imágenes [3,4]. Uno de los campos de investigación es su aplicación al apoyo al diagnóstico con imágenes biomédicas [5].

Un ámbito de aplicación es el apoyo durante un reconocimiento invasivo. *Gastrointestinal Image ANALysis*, GIANA en adelante, es un evento que propone una serie de desafíos, entre los que se encuentra la localización de pólipos durante una colonoscopia [6,7]. Se denomina pólipo al tejido que crece de forma anormal por debajo de la pared del músculo peristáltico y que puede llegar a ser canceroso. Su presencia puede verse reflejada en el tracto del colon como una protuberancia, pero con diversas apariencias. Muchas de estas apariencias son poco apreciables y puede confundirse con otros elementos naturales o patológicos de otra índole (Figura 1).



**Figura 1.** Muestras de pólipo. Se manifiestan como pequeñas protuberancias que pueden pasar desapercibidas (de izquierda a derecha, los dos primeros – en la parte superior central de la imagen y en la inferior central, respectivamente –).

El objetivo de este trabajo fue la selección y creación de una metodología que pueda afrontar el problema planteado por GIANA 2018 para la detección de pólipos empleando un método de DL. El presente documento plasma este trabajo en tres partes. En primer lugar, se definen los datos proporcionados, los métodos de DL considerados para afrontar el problema, así como las técnicas planteadas para la manipulación y adaptación de los datos al modelo. En segundo lugar, se justifica el método escogido, la arquitectura concreta que se va a emplear, las técnicas que se van a aplicar sobre el conjunto de datos y sobre el modelo para entrenarlo. Finalmente, se exponen las conclusiones.

## METODOLOGÍA

### Conjunto de datos

GIANA proporcionó en su edición de 2018 un conjunto de 18 vídeos de colonoscopia. En todos los vídeos se descubre la existencia de un pólipo, y sólo uno. Los pólipos no se visualizan durante toda la duración del vídeo, sino que puede pasar un intervalo de tiempo hasta localizarse y ser enfocado por la cámara, o seguir haciendo un reconocimiento de otras zonas tras haber localizado el pólipo. Los videos se proporcionan segmentados en los fotogramas que lo componen, 11954 imágenes en total. Asociado a cada fotograma se proporciona una máscara, una imagen binaria de misma resolución que marca una región circular donde se encuentra el pólipo si existiera en la imagen.

### Métodos de detección de objetos

Se estableció realizar en una primera fase una investigación para analizar los métodos de DL surgidos en los últimos 3 años (desde inicios de 2015). Se seleccionaría un modelo a utilizar para nuestro problema en base al rendimiento demostrado, la rapidez de ejecución, la facilidad para adaptar el conjunto de datos al modelo y la flexibilidad del método para adaptarlo al problema. Una vez seleccionado el método, seleccionamos una arquitectura ya probada para la parte configurable del método y así establecer un modelo.

### Variaciones en la base de datos y control de parámetros

Se investigaron las modificaciones que pudieran realizarse al conjunto de datos para adaptarlo al modelo establecido, así como para alcanzar un mejor rendimiento durante el entrenamiento. Se establecieron los cambios a probar en la base de datos y se determinaron técnicas a aplicar al modelo con los que se hicieran diferentes entrenamientos del modelo.

## RESULTADOS Y DISCUSIÓN

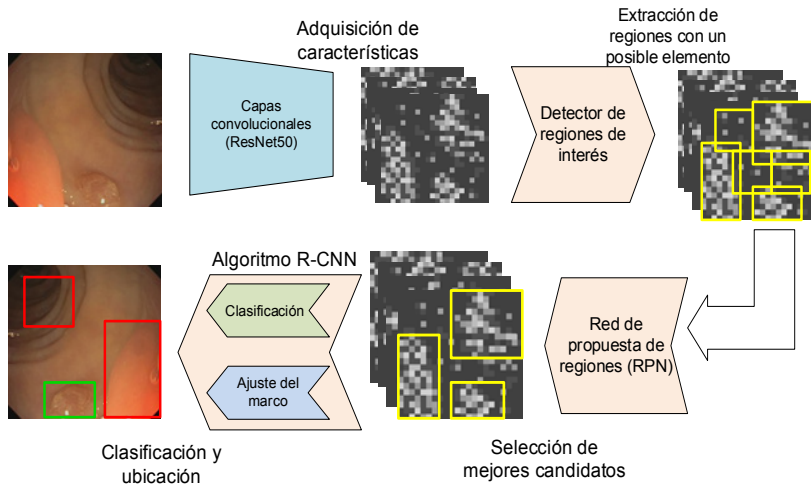
### Análisis de métodos de detección de objetos

Son varios los métodos de DL surgidos en los últimos 3 años para identificar objetos en regiones de imágenes. Se investigaron *Faster Regional Convolutional Neural Network (Faster R-CNN)* [8], *Mask Regional Convolutional Neural Network (Mask R-CNN)* b [9], *Regional Fully Convolutional Network (R-FCN)* [10], y *Single Shot Multibox Detector (SSD)* [11] por ser los más consolidados en la actualidad.

*Faster R-CNN* es un método que surgió con el objetivo de mejorar la eficiencia de su antecesor *Fast R-CNN* mediante el reaprovechamiento del cómputo realizado en la primera etapa del algoritmo para evitar cálculo repetitivo. *Mask R-CNN* es un modelo basado en *Faster R-CNN* cuya mejora reside en la capacidad de devolver un contorneado preciso del objeto detectado. *R-FCN* es un método también basado en *Faster R-CNN* que mejora considerablemente la velocidad al hacer todas sus etapas del algoritmo con capas de convolución. *SSD* es un método similar a los anteriores que se caracteriza por no tener fases separadas en ejecución, sino que se realizan todas a la vez, lo que confiere una eficiencia superior a los otros métodos.

Dado que la capacidad de devolver contornos de *Mask R-CNN* supone un coste computacional mayor y que nuestro objetivo es el de localizar la región donde se localiza cada pólipo en vídeos, este método fue descartado. *Faster-RCNN* supera en rendimiento al resto de los algoritmos que consideramos [12]. A costa de una reducción en la exactitud, *SSD* y *R-FCN* poseen una velocidad mayor, debido al aprovechamiento de la capacidad de cómputo paralelo que de unidades de procesamiento avanzadas. *Faster-RCNN* en cambio posee una flexibilidad mayor al permitir seleccionar el número de regiones que se quiere considerar en el entrenamiento, lo que permite una adaptación mejor a problemas diferentes en los que, como es nuestro caso, no se requiera detectar muchos objetos en una misma imagen.

Se escogió el método *Faster R-CNN* por su flexibilidad a la hora de personalizar cada una de sus etapas, así como la capacidad para especificar la cantidad de regiones que se quiere analizar en cada imagen, que posibilita reducir el coste computacional. Este método, esquematizado en la Figura 2, extrae las características generales de la imagen en una primera fase. Posteriormente decide qué regiones puede contener un objeto o ítem a identificar en base a las características previamente calculadas. Seguidamente emplea una técnica denominada Region Proposal Network (RPN), que va analizando qué regiones de interés -tantas como previamente se hayan establecido- son las mejores candidatas para contener un ítem. Finalmente, una combinación de dos redes neuronales se encarga de clasificar dicho objeto y reajustar la región propuesta para que abarque el objeto de forma más precisa. Esta última fase consiste en la fase característica de su método predecesor *Fast-RCNN* [13]. El modelo escogido a partir de este método emplea una arquitectura *ResNet50* [14] en su primera etapa de extracción de características. Las arquitecturas *ResNet* resuelven el problema de rendimiento que poseen las redes convolucionales que tienen una gran cantidad de capas, introduciendo entre ellas lo que se denomina capas residuales.



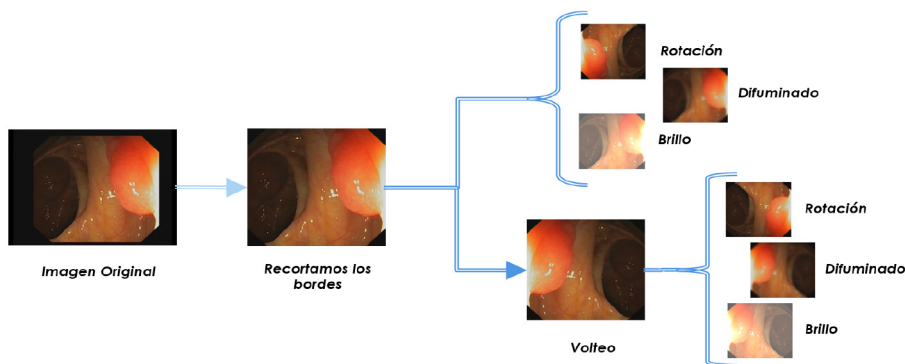
**Figura 2.** Diagrama de la arquitectura de Faster-RCNN empleada con ResNet50 aplicada en la adquisición de características.

## Modificaciones en la base de datos y parámetros

El método usado devuelve regiones rectangulares determinados por coordenadas, por lo que requiere las coordenadas de las imágenes del conjunto de entrenamiento. Se generaron las coordenadas del rectángulo de área mínima que recubriese la región circular proporcionada por cada imagen de los vídeos de colonoscopia.

Las imágenes originales poseen un marco negro que fue eliminado para el entrenamiento del modelo (conjunto A), por lo que se hizo un recalcule de las coordenadas de la región que contiene un pólip.

Habitualmente para aumentar la variedad del conjunto de datos y combatir el sobreajuste del modelo se recurre a técnicas de data augmentation. Consisten en manipulaciones de las imágenes del subconjunto original tales como cambio de orientación o brillo. Para nuestro conjunto de datos se aplicó una secuencia de operaciones para ampliarlo (Figura 2). Se creó una imagen volteada horizontalmente de todo el conjunto A (conjunto B). Sobre los conjuntos A y B se aplicó aleatoriamente una modificación del brillo, un difuminado o una rotación de 180°. Se aplicó una y solo una de estas operaciones por cada imagen. Con este procedimiento se obtuvo un conjunto de datos compuesto por 47816 imágenes diferentes.



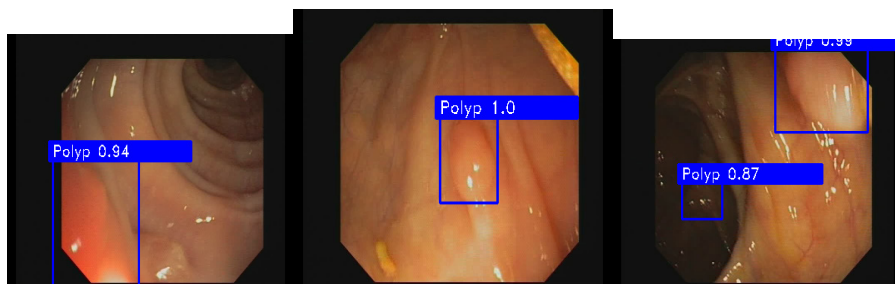
**Figura 3.** Operaciones realizadas sobre el conjunto de datos original.

El algoritmo *Faster R-CNN* en una de sus etapas de entrenamiento aprende a detectar el fondo de la imagen, que se interpreta como un tipo de objeto auxiliar. No obstante, es posible incluir el aprendizaje de este tipo de objetos en las demás etapas proporcionando como entrada regiones que contengan solo fondo. Se ha aplicado esta técnica, denominada *hard negative mining*, utilizando para ello las imágenes que no contienen pólipo.

## CONCLUSIONES

Se ha empleado un algoritmo *Faster R-CNN* para afrontar el problema de detección y ubicación de pólipos en videos de colonoscopia propuesto por GIANA en su edición de 2018. Se ha usado *ResNet50* como arquitectura en la fase de extracción de características de imagen que posee el método. Se han realizado técnicas de *data augmentation* para obtener un conjunto de datos mayor que combata el sobreajuste y mejore el rendimiento del modelo. Se ha empleado *hard negative mining* para reforzar el aprendizaje del fondo y que influya favorablemente en los resultados.

Algunos de los resultados preliminares, en los que se ha empleado parte de los videos de entrenamiento (15 videos de entrenamiento y 3 para evaluación), han alcanzado una precisión del 74.87% y una exactitud de 84.46%. Algunas muestras se ilustran a continuación.



**Figura 4.** Ejemplos sobre el subconjunto de videos empleado en estudios preliminares. A la izquierda se ilustran un caso de falso positivo y un falso negativo (pólipo no identificado en la parte inferior central). En el centro, un caso de verdadero positivo. A la derecha, un caso de falso positivo al fondo y un caso de verdadero positivo.

Los resultados del desafío GIANA serán publicados en el artículo resumen del evento.

## AGRADECIMIENTOS

Agradecemos a los doctores Jorge Bernal y Aymeric Histace su labor como organizadores del *challenge Gastrointestinal Image ANALysis*, así como por proporcionar el conjunto de datos necesario para llevar a cabo nuestra labor investigadora.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Touretzky, D. S., Mozer, M. C., y Hasselmo, M. E. (1996). *Advances in Neural Information Processing Systems 8: Proceedings of the 1995 Conference*, 8
- [2] Kasey, P. (2017). *Top Trends in the Gartner Hype Cycle for Emerging Technologies*. Recuperado de: <https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/>
- [3] LeCun, Y., Bottou, L., Bengio, Y., y Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings IEEE*.
- [4] Krizhevsky, A., Sutskever, I., y Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural information processing systems*, 25(2).
- [5] Greenspan, H., Van Ginneken, B., y Summers, R. M. (2016). Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*.



- [6] Angermann, Q. *et al.* (2017) Towards Real-Time Polyp Detection in Colonoscopy Videos: Adapting Still Frame-Based Methodologies for Video Sequences Analysis. En Cardoso, M., *et al.* (eds). Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures. CARE 2017, CLIP 2017. Lecture Notes in Computer Science, vol 10550. Springer, Cham.
- [7] Bernal, J., *et al.* (2018). Polyp detection benchmark in colonoscopy videos using gtcreeator: A novel fully configurable tool for easy and fast annotation of image databases. Proceedings of 32nd CARS conference, June 2018, Berlin, Germany. (hal-01846141)
- [8] Ren, S., He, K., Girshick, R., y Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [9] He, K., Gkioxari, G., Dollár, P., y Girshick, R. (2017). Mask R-CNN. En *Proceedings of the IEEE International Conference on Computer Vision*.
- [10] Dai, J., Li, Y., He, K., y Sun, J. (2016). R-FCN: Object Detection via Region-based Fully Convolutional Networks. arXiv:1605.06409v2
- [11] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., y Berg, A. C. (2016). SSD: Single shot multibox detector. En *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*,
- [12] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. En *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*.
- [13] Girshick, R. (2015). Fast R-CNN. En *Proceedings of the IEEE International Conference on Computer Vision*.
- [14] He, K., Zhang, X., Ren, S., y Sun, J. (2016). Deep Residual Learning for Image Recognition. *Multimedia Tools and Applications*, arXiv:1512.03385v1.