# Desafíos del aprendizaje profundo en la visión por computador

Víctor Hugo Ayma Quirita (moderador)
vayma@ulima.edu.pe
https://orcid.org/0000-0002-0284-2610
Universidad de Lima, Perú

Pedro Marco Achanccaray Díaz
p.diaz@tu-braunschweig.de
https://orcid.org/0000-0002-7324-9611
Institute of Geodesy and Photogrammetry
Technical University of Braunschweig, Alemania

Smith Washington Arauco Canchumuni
saraucoc@puc-rio.br
https://orcid.org/0000-0003-0812-0085
Pontifical Catholic University of Rio de Janeiro, Brasil

Pedro Juan Soto Vega
pjsotove@ifremer.fr
https://orcid.org/0000-0001-5396-8531
Institut Français de Recherche pour l'Exploitation de la Mer, Francia

Recibido: 3 de septiembre del 2022 / Aceptado: 5 de octubre del 2022
doi: https://doi.org/10.26439/ciis2022.6070

RESUMEN. La visión por computador es un área de estudio en la inteligencia artificial que se enfoca en el desarrollo de técnicas computacionales para percibir el mundo a través de entradas visuales, como videos o imágenes. El aprendizaje profundo ha demostrado ser una técnica eficiente para el análisis e interpretación de datos visuales. Sin embargo, afronta innumerables desafíos según su aplicación en las diferentes tareas de la visión por computador. Este panel reúne un grupo de expertos en aprendizaje profundo, quienes ofrecerán información sobre su aplicación y los desafíos en sus respectivas áreas de investigación con relación a la visión por computador.

PALABRAS CLAVE: visión por computador, aprendizaje profundo

# CHALLENGES OF DEEP LEARNING IN COMPUTER VISION

ABSTRACT. Computer vision is a field of study within artificial intelligence that focuses on developing computational techniques to perceive the world through visual data, such as video or images. Deep learning has proven to be efficient in visual data analysis and interpretation. Nevertheless, it faces countless challenges, given its application in several computer vision tasks. This panel brings together deep learning experts, who will share information about deep learning applications and challenges to overcome in their research fields regarding computer vision.

KEYWORDS: computer vision, deep learning

## 1. INTRODUCCIÓN

La visión por computador es una subárea de la inteligencia artificial que se enfoca en el desarrollo de técnicas computacionales para el análisis e interpretación de datos visuales, como videos o imágenes (Prince, 2012). Se ha convertido en una tecnología fundamental para muchos campos de la industria, como la seguridad, el cuidado de la salud, la agricultura, el entretenimiento, así como la industria textil y automotriz.

El aprendizaje profundo es un conjunto de técnicas basadas en redes neuronales (RN) que han contribuido al desarrollo de la visión por computador. Desde el sorprendente rendimiento alcanzado por las redes neuronales convolucionales (CNN) en la competencia ImageNet[1], estas se han convertido en los modelos de aprendizaje profundo de referencia para la clasificación de imágenes (Yu et al., 2022; Krizhevsky et al., 2012), detección de objetos (Long et al., 2020), segmentación semántica (Zoph et al., 2020) y estimación de la postura humana (Cao et al., 2021). No obstante la popularidad de las CNN, otros modelos de aprendizaje profundo, como los *autoencoders* (AE), las redes neuronales recurrentes (RNN) y las redes generativas adversariales (GAN), han demostrado ser eficientes en la realización de diferentes tareas de la visión por computador. Por ejemplo, las RNN se han aplicado en el reconocimiento de acciones (Singh et al., 2017) y en el de escritura (Carbune et al., 2020); los AE se han aprovechado de forma eficiente para eliminar el ruido de imágenes (Bajaj et al., 2020) y realizar búsquedas en la web con base en imágenes; las GAN se han aplicado en la generación de imágenes realistas a partir de texto y bocetos (Lu et al., 2018; Reed et al., 2016), así como en la generación de vistas frontales de rostros para sistemas de reconocimiento facial.

A pesar de los grandes avances que ha experimentado el aprendizaje profundo en la última década, y los impresionantes resultados alcanzados en la realización de tareas de visión por computador, los diferentes modelos de redes neuronales que componen la tecnología del aprendizaje profundo aún enfrentan desafíos que necesitan ser atendidos.

## 2. PRESENTACIÓN

Este panel reúne un grupo de tres expertos en aprendizaje profundo y visión por computador con reconocida trayectoria en investigación, proyectos de desarrollo e innovación, quienes ofrecerán sus perspectivas sobre los avances en este campo y los desafíos que esta tecnología enfrenta en sus áreas de investigación relacionadas con la visión por computador. En primer lugar, se presentará a los panelistas y se facilitará las preguntas y respuestas del público. Luego se hará un resumen de la sesión, así como una breve introducción al aprendizaje profundo y sus aplicaciones en la teledetección; se comentarán los desafíos que esta tecnología afronta

1 Imagenet: base de datos de imágenes. Disponible en https://image-net.org/

en el análisis de imágenes satelitales para el análisis de cultivos. Posteriormente, la discusión se enfocará en el uso del aprendizaje profundo para la generación de muestras, con especial interés en la generación de datos sintéticos de facies geológicas para el estudio de modelos de reservorios de petróleo. Finalmente, se abordarán los desafíos del aprendizaje profundo en la detección de deforestación en regiones amazónicas a partir del análisis de imágenes de satélite con dicha tecnología.

## 3. CONCLUSIONES

Este foro reunió a un panel conformado por tres expertos en visión por computador y aprendizaje profundo, quienes comentaron sobre los desafíos a los que se enfrentan las técnicas de aprendizaje profundo en el desarrollo de aplicaciones de visión por computador. El panel coincidió en que los mayores desafíos están relacionados con la disponibilidad de bases de datos con información suficiente para la generación de arquitecturas de aprendizaje profundo, así como con la fiabilidad de datos obtenidos a partir de técnicas de generación de muestras sintéticas.

## REFERENCIAS

Bajaj, K., Singh, D. K., & Ansari, M. A. (2020). Autoencoders based deep learner for image denoising. *Procedia Computer Science*, *171*, 1535-1541. https://doi.org/10.1016/j.procs.2020.04.164

Cao, Z., Hidalgo, G., Simon, T., Wei, S., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(01), 172-286. https://doi.org/10.1109/TPAMI.2019.2929257

Carbune, V., Gonnet, P., Deselaers, T., Rowley, H. A., Daryin, A., Calvo, M., Wang, L.-L., Keysers, D., Feuz, S., & Gervais, P. (2020). Fast multi-language LSTM-based online handwriting recognition. *International Journal on Document Analysis and Recognition*, *23*, 89-102. https://doi.org/10.1007/s10032-020-00350-4

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing*, *25*, 1-9.

Long, X., Deng, K., Wang, G., Zhang, Y., Dang, Q., Gao, Y., Shen, H., Ren, J., Han, S., Ding, E., & Wen, S. (2020). PP-YOLO: An effective and efficient implementation of object detector. *ArXiv e-prints*. https://doi.org/10.48550/arXiv.2007.12099

Lu, Y., Wu, S., Tai, Y. W., & Tang, C. K. (2018). Image generation from sketch constraint using contextual GAN. En V. Ferrari, M. Hebert, C. Sminchisescu & Y. Weiss (Eds.),

*Computer Vision – ECCV 2018. ECCV 2018. Lecture notes in computer science* (vol. 11220, pp. 213-228). https://doi.org/10.1007/978-3-030-01270-0_13

Prince, S. (2012). *Computer vision: Models, learning, and inference*. Cambridge University Press.

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. En *Proceedings of the 33rd International Conference on Machine Learning*, *48*, 1060-1069.

Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., & Holzinger, A. (2017). Human activity recognition using recurrent neural networks. En A. Holzinger, P. Kieseberg, A. Tjoa & E. Weippl (Eds.), *Machine Learning and Knowledge Extraction. CD-MAKE 2017. Lecture notes in computer science* (vol. 10410, pp. 267-274). https://doi.org/10.1007/978-3-319-66808-6_18

Yu, J., Wang, Z., Vasudevan, V., Yeung, L., Seyedhosseini, M., & Wu, Y. (2022). CoCa: Contrastive captioners are image-text foundation models. *ArXiv e-prints*. https://doi.org/10.48550/arXiv.2205.01917

Zoph, B., Ghiasi, G., Lin, TY., Cui, Y., Liu, H., Cubuk, E. D., & Le, Q. (2020). Rethinking pre-training and self-training. *Advances in Neural Information Processing*, *33*, 1-13.