

Análisis bibliométrico de la aplicación de redes neuronales convolucionales en el área de la visión artificial

Bibliometric analysis of the application of convolutional neural networks in the area of computer vision

Marco A. De la Cruz-Rocca^{1,*}, Williams Muñoz-Robles^{1,#}, Melquiades A. Trinidad-Malpartida^{1,%}

Resumen

Este artículo analiza la aplicación de las Redes Neuronales Convolucionales (RNC) en el campo de la visión artificial, mediante el método bibliométrico. Se realiza un análisis de muestras literarias, utilizando estadísticas descriptivas básica, mediante un filtro de la base de datos Scopus de 2,526 registros, comprendido en un periodo de 5 años (2018-2022). La revisión del marco teórico revela que las RNC se emplean en diversas aplicaciones de visión artificial como reconocimiento de imágenes, clasificación de objetos, detección de patrones y otras aplicaciones relacionadas con el procesamiento de imágenes incluyendo el diagnóstico de fallas. El análisis bibliométrico revela un notable incremento en la producción de artículos sobre RNC en el ámbito de la visión artificial, abarcando diversas disciplinas, con una alta concentración de documentos en ciencias de la computación e ingeniería. Las organizaciones chinas destacan por su alta proporción de afiliaciones y capacidad de financiamiento de investigaciones en este tema, consolidando a China como líder en la cantidad de publicaciones sobre RNC. El investigador más destacado es A.W. Schumann, aunque no se mantiene consistentemente como el autor principal en otros estudios similares. Las direcciones clave para investigaciones futuras incluyen la exploración experimental cuantitativa y la diversificación de los campos de acción, ampliando así el impacto y la aplicación de las RNC en nuevas áreas.

Palabras clave: redes neuronales convolucionales, visión artificial, análisis bibliométrico.

Abstract

This article analyzes the application of Convolutional Neural Networks (CNN) in the field of computer vision, using the bibliometric method. An analysis of literary samples, using basic descriptive statistics, is performed by filtering the Scopus database of 2526 records, comprising a period of 5 years (2018-2022). The review of the theoretical framework reveals that CNRs are employed in various computer vision applications such as image recognition, object classification, pattern detection and other applications related to image processing including fault diagnosis. The bibliometric analysis reveals a notable increase in the production of articles on RNC in the field of computer vision, covering various disciplines, with a high concentration of documents in computer science and engineering. Chinese organizations stand out for their high proportion of affiliations and capacity to finance research on this topic, consolidating China as the leader in the number of publications on RNC. The most prominent researcher is A.W. Schumann, although he is not consistently maintained as the lead author in other similar studies. Key directions for future research include quantitative experimental exploration and diversification of fields of action, thereby expanding the impact and application of RNCs in new areas.

Keywords: convolutional neural networks, computer vision, bibliometric analysis.

¹Universidad Nacional Daniel Alcides Carrión, Cerro de Pasco, Perú

E-mail: *macruz@undac.edu.pe, #wmunozr@undac.edu.pe, %atrinidadm@undac.edu.pe

Orcid ID: ^a<https://orcid.org/0000-0003-3681-7963>, ^b<https://orcid.org/0000-0002-8896-1724>, ^c<https://orcid.org/0000-0001-6286-155X>

Recibido: 30 de diciembre de 2023

Aceptado para publicación: 20 de febrero de 2024

Publicado: 30 de marzo de 2024

Citar este artículo: De la Cruz-Rocca, M.A., Muñoz-Robles, W., Trinidad-Malpartida, M.A. (2024). Análisis bibliométrico de la aplicación de redes neuronales convolucionales en el área de la visión artificial. *Investigación Valdizana*, 18(1), e2105.
<https://doi.org/10.33554/riv.18.1.2105>

Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0)



Introducción

La investigación científica en la actualidad sigue siendo una actividad esencial en la sociedad moderna. La necesidad de entender y explicar los fenómenos naturales y sociales, desarrollar nuevas tecnologías y abordar problemas globales como el cambio climático, la salud pública y la justicia social, son algunos de los campos de acción que impulsan la investigación científica.

Un campo muy interesante que ha tenido un acelerado avance científico dentro de la ingeniería es la visión artificial y con ella las redes neuronales convolucionales (RNC), que son un tipo de red neuronal artificial utilizado ampliamente en el campo de la visión artificial debido a su capacidad para extraer características y patrones útiles de imágenes y videos, permitiendo reconocer objetos, identificar personas, optimizar la seguridad personal y organizacional, detectar anomalías de calidad en productos, entre muchas otras cosas.

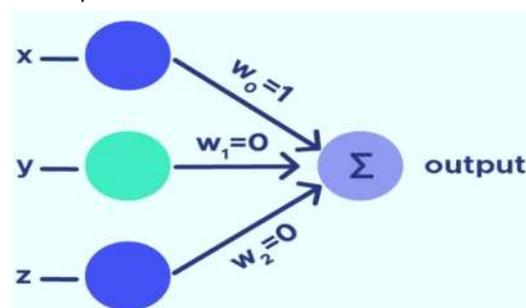
Estas redes, inspiradas en el funcionamiento del sistema visual biológico, han demostrado eficacia en tareas específicas de procesamiento de imágenes y reconocimiento de patrones. No obstante, a pesar de los avances significativos, persisten desafíos y preguntas no resueltas en relación con la aplicación de las redes neuronales convolucionales en el campo de la visión artificial. La complejidad inherente a estos modelos, así como su capacidad para adaptarse y aprender de grandes conjuntos de datos, plantea interrogantes sobre su rendimiento en distintos escenarios.

En este contexto, surge la necesidad de un análisis de la aplicación de las redes neuronales convolucionales en el ámbito de la visión artificial que permita responder a la siguiente interrogante ¿Cuál es el estado y la tendencia de la aplicación de las redes neuronales convolucionales para resolver problemas en visión artificial?, la formulación de este problema busca abordar las incertidumbres y limitaciones existentes sobre este particular, en ese sentido y con el fin de obtener un norte claro en la investigación nos planteamos el siguiente objetivo general: Analizar el estado y la tendencia de la aplicación de las redes neuronales convolucionales para resolver problemas en el área de la visión artificial. Para tal fin se plantea una serie de objetivos específicos: 1) Realizar una revisión bibliográfica sobre redes neuronales convolucionales y visión artificial en un periodo de 5 años anteriores; 2) Realizar un análisis bibliométrico de la acción de las redes neuronales convolucionales para resolver problemas en visión artificial; y 3) Describir las tendencias de las redes neuronales artificiales convolucionales. La comprensión precisa de estos aspectos no solo contribuirá al desarrollo continuo de la visión artificial, sino que también permitirá aprovechar al máximo el potencial de las redes neuronales convolucionales en aplicaciones prácticas y diversas, desde el reconocimiento de objetos hasta el análisis de imágenes.

La evolución de la visión artificial y las redes

neuronales convolucionales ha sido marcada por avances significativos en las últimas décadas, impulsados por la intersección de la informática, la inteligencia artificial y la investigación en neurociencia computacional. Los albores de esta disciplina van por la década de 1960, Frank Rosenblatt desarrolló el perceptrón (ver figura 1) dando paso a los primeros modelos de reconocimiento de patrones sentando las bases para la investigación en redes neuronales, sin embargo, a pesar del entusiasmo inicial, la falta de avances significativos y la percepción de limitaciones en los perceptrones llevaron a un estancamiento en la investigación en redes neuronales durante la década de 1980 (Steemit, 2019).

Figura 1
Perceptrón de 3 entradas



Nota: estructura básica de las redes neuronales convolucionales. Fuente: Datascientest (2022).

En la década de los 90 resurge y aparece en escena las Redes Neuronales Convolucionales (CNN), un tipo de red neuronal artificial que se utiliza principalmente para el procesamiento de imágenes y reconocimiento de patrones mediante matrices (García, 2019), Yann Lecun, uno de los pioneros, contribuyó al desarrollo de LeNet-5, una RNC utilizada con éxito en el reconocimiento de dígitos escritos a mano. Esta red se utilizó en aplicaciones prácticas, como la lectura automática de códigos postales en el Servicio Postal de los Estados Unidos (Durán, 2017).

Aunque las CNN mostraron promesa en tareas específicas, durante un tiempo perdieron popularidad debido a la falta de grandes conjuntos de datos y poder computacional suficiente para entrenar modelos más grandes y complejos. El renacimiento de las CNN se produjo con la participación en el concurso ImageNet Large Scale Visual Recognition Challenge (ILSVRC) en 2012. Alex Krizhevsky, junto con otros investigadores, presentó la arquitectura AlexNet, una CNN profunda que superó significativamente a los métodos tradicionales y desencadenó un nuevo interés en las redes neuronales convolucionales (Durán, 2017).

Desde entonces, ha habido numerosos avances en arquitecturas de RNC, como VGG (red neuronal convolucional con 19 capas de profundidad), GoogLeNet, ResNet, etc. Estas redes han demostrado ser efectivas en una amplia gama de tareas de visión por computadora, como clasificación de imágenes, detección de objetos, segmentación semántica, entre otras (Steemit, 2019).

Hoy en día el accionar de esta tecnología RNC se ha extendido a muchos campos donde se trabaje con visión artificial: robótica, medicina, minería, etc, siendo por ello necesario hacer un análisis de cómo opera, lo que ha logrado y tendencias de esta tecnología.

Revisión bibliográfica

Antecedentes de la investigación

En las últimas décadas, el aumento exponencial en la disponibilidad de datos y el desarrollo de tecnologías de procesamiento han impulsado avances significativos en el campo de las RNC siendo su campo de acción principal la visión artificial para el reconocimiento de imágenes, tanto en objetos inertes como vivos. Entender cuanto se ha hecho en este tema involucra una revisión exhaustiva de artículos bibliométricos y revisiones del tema. A continuación, se mencionan los más relevantes.

Barcic et al. (2023) en su artículo científico basado en una revisión de 150 artículos, ofrece valiosos datos sobre diversos aspectos del reconocimiento facial basado en RNC. Se destaca que las conferencias son la principal plataforma de publicación, seguidas por las revistas, mientras que los talleres y simposios tienen una presencia más limitada. El análisis temporal revela un aumento significativo en la cantidad de artículos publicados de 2013 a 2017, indicando un creciente interés y avance en el campo. Se produce una estabilización de publicaciones de 2018 a inicios de 2020, tras lo cual se nota un repunte en lo que sigue del 2020, lo que sugiere una dinámica de investigación en evolución, posiblemente en respuesta a tecnologías emergentes. En cuanto a las arquitecturas CNN, VGGNet, ResNet y AlexNet dominan, sirviendo como bases sólidas y puntos de referencia en el campo del reconocimiento facial. Otras arquitecturas como GoogLeNet, Inception, MobileNet y DenseNet también han contribuido significativamente al avance de este dominio. Durante el periodo intermedio de desarrollo de las RNC entre 2017 y 2020, el enfoque predominante se centró en el aprendizaje profundo (Deep learning). Este enfoque estuvo marcado por una atención especial a la mejora del rendimiento, el uso de conjuntos de datos a gran escala, el énfasis en el reconocimiento en tiempo real y el empleo del aprendizaje por transferencia.

Así mismo Chen & Deng (2020) realizaron un estudio donde examina el avance de la investigación en el ámbito de las RNC mediante el empleo del enfoque bibliométrico en la base de datos de la Web of Science, para un total de 4,598 artículos, durante 20 años. Se realiza un análisis de la literatura relacionada con las RNC a través de estadísticas básicas, indican que en relación a las citas de artículos de RNC sólo tenía 78 citas en 2010, pero 60,739 en 2020, esto revela un crecimiento exponencial. Entre las muestras de literatura, la más citada es el estudio realizado por Krizhevsky et al. (2012), por otro lado, los principales campos de investigación de la bibliografía sobre las RNC son la ingeniería, la informática, la radiología, la medicina nuclear, el diagnóstico por imagen y las telecomunicaciones. También indican que los trabajos más publicados en todas las muestras bibliográficas proceden de la Academia

China de Ciencias con 246 trabajos publicados, consecuentemente de las diez instituciones de investigación con más artículos publicados, nueve están en China y una en Estados Unidos. De otra parte, se observa que los 45 autores principales publicaron un total de 42 artículos, lo que representa el 22.22% del total de artículos publicados. De estos autores, Sun X y Zhang Y publicaron nueve artículos cada uno. En cuanto a número de citas de autor, indican que Frizhevsky encabeza las citas por publicación con 1,347 por su artículo.

“ImageNet classification with deep CNNs”, dentro de sus conclusiones los autores indican que los resultados experimentales revelan que las RNC están siendo aplicadas en diversas áreas de la visión artificial, abarcando desde el diagnóstico de fallos y el reconocimiento de imágenes hasta la detección sísmica, posicionamiento, detección automática de grietas y señales, clasificación y segmentación de imágenes. Se anticipa que la investigación cuantitativa experimental, la expansión de los campos de aplicación y la informatización de los estudios de mercado serán las tendencias cruciales en la investigación futura.

Prahani et al. (2023) indica que sin lugar a dudas las empresas de hoy en día están recurriendo cada vez más a algoritmos de aprendizaje. Estas tecnologías están comúnmente asociadas con la Inteligencia Artificial (IA), el Aprendizaje Profundo (AP), el Aprendizaje Automático (ML) y las Redes Neuronales Artificiales (RNA), dentro de la cual se encuentra las RNC. Esta investigación realiza un análisis bibliométrico que permite recabar información descriptiva sobre el particular, basado en 809,045 documentos. Indica también que en el año 2022 se produjo el mayor número de publicaciones en todas las áreas mencionadas, con tendencia a seguir creciendo. En cuanto a las principales áreas de investigación en IA, ML, DL y RNA menciona que ingeniería y ciencia de datos ocupan las primeras posiciones, en ese orden, en cuanto a instituciones investigativas la Academia China de Ciencias es la más importante tanto en ponencias como en artículos.

Finalmente, este estudio indica que se espera un aumento anual en la publicación de investigaciones sobre IA, ML, AP y RNA ya que un análisis del crecimiento de las publicaciones en estos temas en la última década indica un aumento constante cada año.

Aloysius & Geetha (2017) mencionan que las RNC han surgido como una alternativa valiosa para el aprendizaje automático de características específicas en diversos dominios, siendo el más llamativo la visión por computadora, utilizando AlexNet como modelo base de RNC frente a otros seis modelos, siendo AlexNet la que mejores resultados arroja, sin embargo, indica que no se puede explicar teóricamente por qué funcionan tan bien.

Bases conceptuales

Visión Artificial

La visión artificial, también conocida como visión informática o visión por computadora (del inglés computer



vision) o visión técnica (García y Caranqui, 2015) es definida por Borrella (2022) como “una disciplina científica que incluye métodos para adquirir, procesar y analizar imágenes del mundo real con el fin de producir información que pueda ser tratada por una máquina”, así mismo Forsyth y Ponce (2002) en su publicación dan la siguiente definición “La visión por computadora es el estudio de cómo enseñarle a las computadoras a ver” (p. 181), con el propósito de equipar a las computadoras con habilidades visuales análogas a las humanas, permitiéndoles interpretar, comprender y responder a imágenes y secuencias visuales del mundo real.

Como se observa ambos autores abordan la definición y el propósito de la visión por computadora desde perspectivas complementarias. La primera descripción se enfoca en la disciplina científica en sí misma, presentando la visión por computadora como un campo que abarca métodos para adquirir, procesar y analizar imágenes del mundo real con el objetivo de generar información procesable por máquinas. Esta definición destaca la naturaleza técnica y metodológica de la disciplina, subrayando su aplicación práctica para extraer información útil de datos visuales. Por otro lado, la segunda descripción va más allá al conceptualizar la visión por computadora como el estudio de cómo enseñar a las computadoras a “ver”. Esta perspectiva resalta el objetivo de equipar a las máquinas con habilidades visuales comparables a las humanas, subrayando la capacidad de interpretar, comprender y responder a imágenes y secuencias visuales del mundo real. Aquí, se destaca la aspiración de lograr una percepción visual en las computadoras que se asemeje a la capacidad humana, apuntando hacia una comprensión más profunda e interactiva del entorno visual.

Tanto en un dispositivo informático como en un organismo animal, la visión se compone esencialmente de dos partes. En primer lugar, un sensor capta todos los detalles de una imagen: el ojo recoge la luz a través del iris y la proyecta en la retina, donde células especializadas transmiten la información al cerebro mediante las neuronas. De manera similar, una cámara captura imágenes y transmite píxeles al ordenador. En esta fase, las cámaras superan a los humanos al poder ver en infrarrojo, a mayores distancias o con mayor precisión. En segundo lugar, el dispositivo de interpretación debe procesar la información y extraer su significado. El cerebro humano resuelve esta tarea en múltiples etapas y en diversas regiones cerebrales. La visión por computadora aún se encuentra rezagada respecto al desempeño humano en este ámbito (He et al., 2015).

Las cámaras son omnipresentes, y la cantidad de imágenes cargadas en Internet crece de manera exponencial. Desde fotos en Instagram hasta videos en YouTube, pasando por instantáneas de cámaras de seguridad e imágenes médicas y científicas, la visión por computadora se vuelve esencial. Su función crucial radica en clasificar estas imágenes y permitir que las computadoras comprendan su contenido. A continuación, se presenta una lista no exhaustiva de aplicaciones de la visión por computadora.

- Efectos especiales

- Modelado 3D
- Reconocimiento de escenas
- Detección facial
- Reconocimiento de texto
- Búsqueda visual móvil
- Coches autónomos
- Realidad aumentada

La visión artificial, un campo fascinante de la inteligencia artificial, se centra en dotar a las máquinas con la capacidad de interpretar y comprender información visual del mundo que nos rodea. Una pieza fundamental en este terreno son las Redes Neuronales Convolucionales, que han revolucionado la forma en que las computadoras procesan y analizan imágenes. La sinergia entre visión artificial y RNC ha transformado sectores como la medicina, la industria automotriz, la seguridad y más. Desde diagnósticos médicos basados en imágenes hasta vehículos autónomos capaces de interpretar su entorno, la combinación de visión artificial y RNC está dando forma a un futuro donde las máquinas no solo “ven” el mundo, sino que también lo comprenden de manera sofisticada y adaptable.

Redes neuronales convolucionales

En el contexto de las RNC y su aplicación a imágenes de alta resolución, surge un desafío significativo. Las redes neuronales convencionales, al ser completamente conectadas, enfrentan problemas de eficiencia cuando se trabaja con imágenes de resoluciones elevadas, como el estándar Full HD (1,980 x 1,080 píxeles). Este problema radica en el elevado número de conexiones necesarias entre las neuronas de la primera capa oculta y cada píxel de la imagen de entrada, lo que resulta en tiempos de entrenamiento y testeo considerablemente prolongados (Durán, 2017).

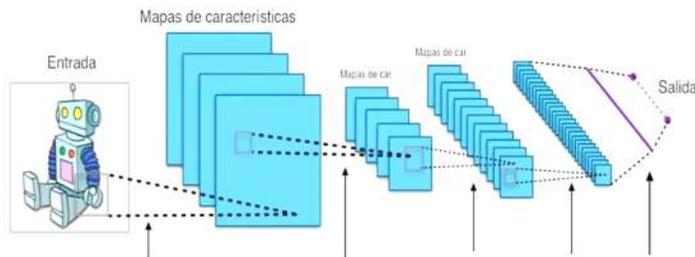
Durán (2017) en su trabajo de investigación define una Red Neuronal Convolutiva como un tipo de arquitectura de red multicapa que incluye capas convolucionales (extracción de características) y de pooling (submuestreo) de manera alternada, seguidas por capas totalmente conectadas similares a una red perceptrón multicapa.

Las RNC constituyen una categoría de redes de aprendizaje profundo, inicialmente diseñadas para el análisis de datos de imágenes. No obstante, también se emplean en el procesamiento del lenguaje natural, detección de anomalías, análisis de vídeo, descubrimiento de fármacos, evaluación de riesgos para la salud, previsión de series temporales y sistemas de recomendación (Glassner, 2021, citado en Barcic et al., 2023).

La entrada típica a una capa convolutiva es una imagen de tamaño $s \times s \times r$, donde s representa tanto la altura como el ancho de la imagen, y r es el número de canales ($r=1$ representa escala de grises). Estas capas convolucionales incorporan filtros (o núcleos) con dimensiones $n \times n \times q$, donde n y q son seleccionados por el diseñador, con q generalmente igual a r . Cada filtro realiza una convolución, generando un mapa de “p”

características de tamaño $(s-n+1) \times (s-n+1)$, siendo p el número de filtros a usar. Luego cada mapa es submuestreado en la capa pooling, para ello puede emplear "mean pooling" o "max pooling", esto genera reducción de dimensión de cada característica filtrada, observe toda esta descripción en la figura 2 (Durán, 2017).

Figura 2
Esquema de una red neuronal convolucional



Nota. Demostración con datos bidimensionales y representaciones ocultas. Fuente: Duran (2017).

a. Capa de convolución

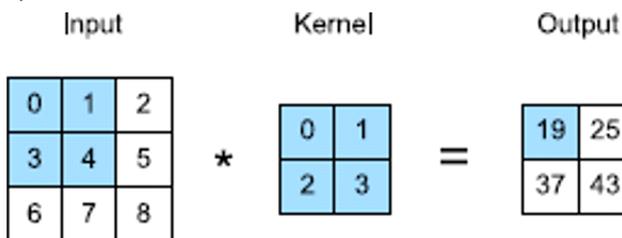
En una capa de este tipo, un tensor (matriz bidimensional) de entrada y un tensor de núcleo se combinan para producir un tensor de salida a través de una operación de correlación cruzada (Zhang et al., 2021).

Ejemplificando esto con datos bidimensionales y representaciones ocultas. En la Figura 3, la entrada es un tensor Input bidimensional (matriz 3 x 3), representa una imagen pixelada, donde cada posición i, j de la matriz contiene un valor numérico de color de un pixel. La altura y la anchura del núcleo (kernel) es 2 (matriz 2x2), representa al filtro que se desea aplicar a la imagen para extraer alguna característica importante de este como bordes, tonos, sombras, líneas, curvas, etc., para ello se asigna valores a cada posición de esta matriz.

Dada estas matrices se procede a multiplicar los valores sombreados de ambas (matriz Input por matriz Kernel), el resultado se ubica en la matriz Output (posición sombreada).
 $0 \times 0 + 1 \times 1 + 3 \times 2 + 4 \times 3 = 19$.

Se procede del mismo modo para obtener los demás valores de esta matriz de salida.

Figura 3
Operación de correlación cruzada bidimensional.



Nota. Demostración con datos bidimensionales y representaciones ocultas. Fuente: Zhang et al. (2023).

En la imagen de la figura 4, Ringa Tech (2021) muestra el resultado en un video de YouTube sobre la

aplicación de una convolución en a una imagen mediante un filtro.

Figura 4
Convolución de una imagen para detectar bordes



Nota. Imagen obtenida de un mapa de características convolucionadas donde se procede a aplicar submuestreo. Fuente: Ringa Tech (2019).

b. Capa de submuestreo o agrupación

Una vez que se ha obtenido un mapa de características convolucionadas se procede a aplicar submuestreo (agrupación o pooling) mediante la capa correspondiente. La capa de submuestreo en una RNC tiene como objetivo reducir la dimensionalidad espacial de la entrada y disminuir el número de parámetros en la red, al mismo tiempo que conserva la información más relevante. La operación de pooling se realiza aplicando una función (como "max pooling" o "mean pooling") sobre regiones locales de la entrada (Krizhevsky et al., 2012).

El resultado de esta reducción se puede apreciar en la figura 5 donde se reducen las imágenes en escala.

Figura 5
Aplicación de Max Pooling en la capa de submuestreo.

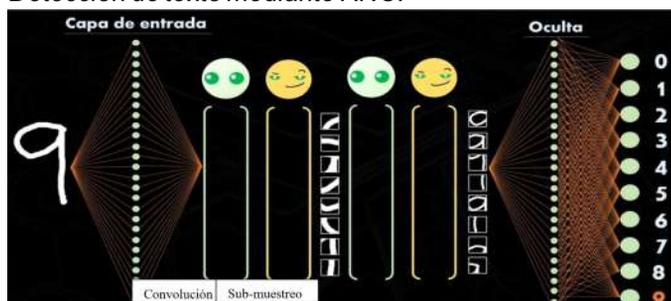


Nota. En el proceso se pueden detectar y reconocer objetos a diferente escala en imágenes. Fuente: Ringa Tech (2019).

Mediante este proceso se pueden detectar y reconocer diferentes objetos en imágenes, como se observa en la figura 6, donde se muestra la secuencia de detección de texto, en este caso el número 9, obsérvese que puede pasar por varias capas de convolución y submuestreo para poder ser reconocido.

Figura 6

Detección de texto mediante RNC.



Nota. Las arquitecturas comunes en RNC han demostrado ser efectivas en diversas tareas de visión por computadora. Fuente: Ringa Tech (2019).

Arquitecturas comunes en RNC

Existen varias arquitecturas comunes en RNC que han demostrado ser efectivas en diversas tareas de visión por computadora. A continuación, se menciona algunas de las arquitecturas más destacadas:

- a. **LeNet-5:** Desarrollada por Yann Lecun en la década de 1990, fue una de las primeras CNN y se utilizó para el reconocimiento de dígitos escritos a mano (Lecun et al., 1998).
- b. **AlexNet:** Presentada por Alex Krizhevsky, Ilya Sutskever y Geoffrey Hinton en 2012, fue la primera CNN en ganar la competición ImageNet Large Scale Visual Recognition Challenge, marcando un avance significativo en la capacidad de las CNN. La red era similar a LeNet, pero en lugar de alternar capas convolucionales y capas agrupadas, AlexNet tenía todas las capas convolucionales apiladas. Y en comparación con LeNet, esta red es mucho más grande y profunda (Krizhevsky et al., 2012).
- c. **VGGNet:** Desarrollada por el Visual Geometry Group (VGG) de la Universidad de Oxford, se caracteriza por su simplicidad y consistencia en la arquitectura, utilizando capas convolucionales de tamaño pequeño, filtros de convolución 3x3 en todas las capas, pero de gran profundidad (Simonyan & Zisserman, 2014).
- d. **GoogLeNet (Inception):** Introducida por Google en 2014, esta arquitectura utiliza módulos de "Inception" que combinan diferentes tamaños de filtros convolucionales en paralelo, permitiendo la extracción de características a diferentes escalas, paso por 3 versiones, siendo la actual Inception-v3 (Szegedy et al., 2015).
- e. **ResNet (Residual Network):** Propuesta por Kaiming He, Xiangyu Zhang, Shaoqing Ren y Jian Sun en 2015, introduce el concepto de conexiones residuales, lo que facilita el entrenamiento de redes muy profundas al mitigar el problema de desvanecimiento del gradiente. El principal inconveniente de este tipo de arquitectura RNC es que resulta muy costosa de evaluar debido al enorme número de parámetros. Sin embargo, el número de parámetros puede reducirse hasta cierto punto eliminando la primera capa totalmente conectada (ya que la mayoría de los parámetros se deben a esta capa), sin que ello afecte al rendimiento (He et al., 2015).
- f. **MobileNet:** Diseñada para aplicaciones en dispositivos móviles, esta arquitectura utiliza operaciones de

convolución separable en profundidad y en anchura para lograr una eficiencia computacional significativa (Howard et al., 2017).

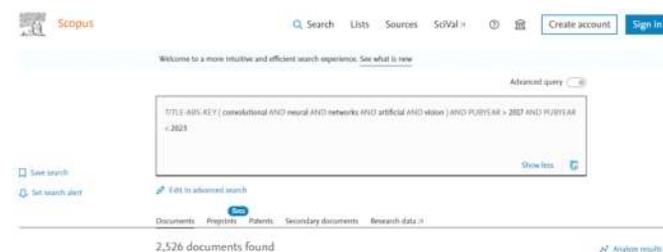
- g. **DenseNet:** Introducida por Gao Huang, Zhuang Liu, y Laurens van der Maaten en 2017, propone conexiones densas entre las capas, donde cada capa recibe directamente las salidas de todas las capas anteriores (Huang et al., 2016).
- h. **EfficientNet:** Desarrollada por Mingxing Tan y Quoc V. Le en 2019, busca optimizar simultáneamente el ancho, la profundidad y la resolución de la red para lograr una eficiencia superior en términos de rendimiento y recursos (Tan & Le, 2020).

Metodología

Esta investigación utiliza análisis descriptivo y bibliométrico (Tomás y Tomás, 2018), para ello se requiere de una base de datos organizada para evaluar los datos de publicaciones, para tal fin se recurrió a Scopus, ya que ofrece el repositorio más centralizado, numerosos temas con variedad de información (Pranckutė, 2021) y por la facilidad de acceso a la misma. Los datos se recopilaron en el mes de junio y julio, con la palabra clave "Convolutional Neural Network" y "artificial vision" en el título de la base de datos central del sitio web Scopus, conforme se puede apreciar en la figura 7.

Figura 7

Criterio de búsqueda de bibliografía en la web Scopus



Nota. Ecuación matemática de criterio de búsqueda de información aplicada en la plataforma de Scopus. Fuente: Elaboración propia con base en web Scopus año 2023.

El ingreso de esta información arrojó en un inicio más de cuatro mil resultados, por lo que se tuvo que mejorar el criterio de búsqueda, circunscribiéndolo a investigaciones al periodo de tiempo comprendido entre los años 2018 y 2022 inclusive, quedando el criterio de búsqueda con 2526 registros listados, siendo el criterio de búsqueda como sigue:

TITLE-ABS-KEY (convolutional AND neural AND networks AND artificial AND vision) AND PUBYEAR > 2017 AND PUBYEAR < 2023

El estudio abordó un quinquenio de literatura internacional, empleando el software Excel para la recopilación de la información obtenida de Scopus.

Análisis de datos y discusión

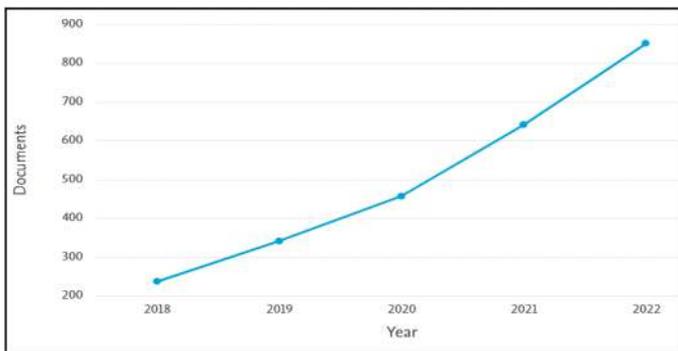
La evaluación del estado de publicación de los artículos se considera comúnmente un indicador crucial para medir el desarrollo de una disciplina y el nivel de logros y contribuciones en la investigación científica. La

RNC destaca como el algoritmo más significativo en el ámbito de la visión por computadora. Esta emplea una red neuronal profunda para emular el proceso de reconocimiento de patrones humano y la cognición del entorno, extrayendo, reconociendo y clasificando características de imágenes. En esta sección, se analizan las tendencias de investigación relacionadas con la RNC mediante estadísticas de publicaciones en revistas y la frecuencia de citas a lo largo de un periodo de tiempo.

Análisis de la tendencia global de crecimiento en publicaciones

La Figura 8 presenta el número de artículos publicados en un periodo de 5 años, destacando un notable incremento en la bibliografía sobre la RCN a partir de 2020 con 457 artículos. El punto más elevado de la producción de artículos científicos se registró en el último año de análisis (2022), con un total de 851, esto representa un incremento de artículos publicados de un 360% en comparación a la cantidad de artículos publicados en el año 2018, que fue de 236.

Figura 8
Tendencias de la investigación en RNC durante cinco años



Nota. La gráfica muestra de la tendencia RCN de artículos publicados en un periodo de 5 años. Fuente: Web Scopus año 2023.

Así mismo, encontramos que el artículo más citado del año 2022 corresponde a Wang et al. con 167 citas. En este documento proponen un decodificador basado en Transformer, construido y denominado UNet (UNetFormer) para la segmentación de escenas urbanas en tiempo real (Wang et al., 2022).

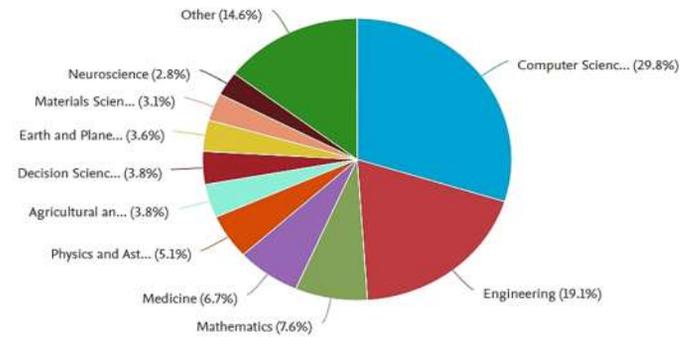
Se puede apreciar la tendencia de crecimiento de la investigación de RNC, esto corrobora lo dicho por Barcic et al. (2023), Prahani et al. (2023) y Chen y Deng (2020), siendo que probablemente esté llegando a la etapa de madurez. Con el auge del aprendizaje profundo, existe una necesidad urgente de comprender y aplicar las RNC en diversos campos, debiendo incrementarse la experimentación para innovar en este tipo de red neuronal, conforme lo indica también Chen y Deng (2020).

Análisis de documentos por área de investigación

La figura 9 revela que los campos principales de investigación en la bibliografía sobre RNC abarcan áreas

como ciencias de la computación, ingeniería, matemática, medicina, física, agricultura entre otros. Siendo ciencias de la computación el área con mayor porcentaje de investigaciones, un 29.8% equivalente a 1,621 documentos, seguida de ingeniería con 19.1% para estos cinco años de análisis.

Figura 9
Principales áreas de investigación y sus proporciones



Nota: La gráfica muestra las proporciones en las áreas de investigación en la bibliografía sobre RNC. Fuente: Web Scopus año 2023.

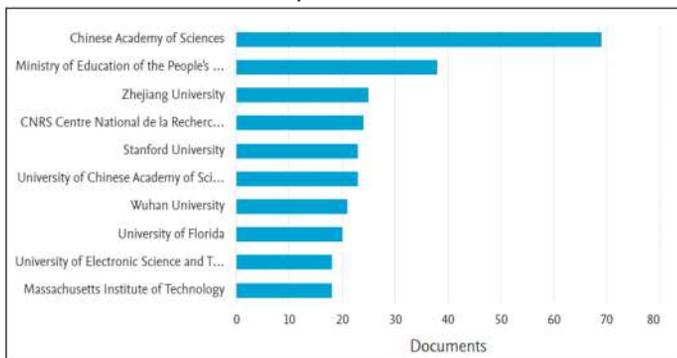
El documento más citado en ciencias de la computación corresponde a Voulodimos et al. (2018) con 1,914 citas, esta investigación realiza una revisión proporcionando un breve resumen de algunos de los esquemas de aprendizaje profundo más destacados empleados en los desafíos de visión por computadora.

Este resultado diverge ligeramente al obtenido por Chen y Deng (2020), ya que en su estudio indican que el área de ingeniería es la que presenta más documentos investigativos, colocando en segundo lugar a la ciencia de la computación, sin embargo, Prahani et al. (2023) si coincide con nuestra investigación en este punto.

Análisis de documentos por filiación a instituciones de investigación científica de alto nivel

Acerca de los institutos de investigación con una destacada producción en el ámbito de la RNC, detallando sus principales áreas de investigación y colaboraciones se presenta la figura 10, donde se observa que la Academia China de Ciencias lidera con el mayor número de artículos publicados en todas las muestras bibliográficas, totalizando 69, a continuación se ubica el Ministerio de Educación de la República Popular China con 38 publicaciones y en un tercer lugar Universidad Zhejiang con 25 artículos publicados.

Figura 10
Cantidad de documentos por filiación institucional



Nota. La tabla muestra la cantidad de artículos publicados de investigación en el entorno RNC. Fuente: Web Scopus año 2023.

De las 10 instituciones de investigación con mayor cantidad de artículos publicados, siete están localizadas en China, dos en Estados Unidos y una en Francia. Estos resultados indican un notable avance en el nivel de investigación sobre la CNN en China en los últimos años. El documento más citado sobre RNC y visión artificial en la Academia China de Ciencias fue publicado por (Xu et al., 2021) en el año 2021, en este artículo, los investigadores presentan un acelerador convolucional vectorial óptico universal que opera a más de diez tera-ops por segundo (TOPS), lo que equivale a billones (10^{12}) de operaciones por segundo.

Finalmente, podemos decir que este análisis de instituciones de investigación científica de alto nivel que considera a la Academia China de ciencias con la primera en el ranking coincide con el resultado obtenido por las investigaciones de Prahani et al. y Chen y Deng. Complementariamente, en la Tabla 1 se muestra la lista de patrocinadores financieros para los trabajos de investigación, siendo la National Natural Science Foundation of China quien ocupa la primera posición con 289 investigaciones financiadas, ocupando el segundo lugar y bastante alejada de la primera, la National Key Research and Development Program of China con 72, la investigación bibliométrica realizada por Prahani et al. señala también a la National Natural Science Foundation of China como la que más proyectos de investigación ha financiado, coincidiendo.

Análisis de documentos por autor principal

Los autores principales son aquellos investigadores que generan una mayor cantidad de literatura publicada y ejercen una mayor influencia en un campo de investigación específica. La ley de Price en bibliometría puede utilizarse para determinar el autor principal en un campo de investigación (Chen y Deng, 2020). Utilizando datos de Scopus, se muestra en la Tabla 2 el número de publicaciones de los diez autores que encabezan la lista, esta revela que Schumann, A.W., ocupa la primera posición con nueve publicaciones, siendo su artículo Deeplearning for image-based weed detection in turfgrass (Aprendizaje profundo para la detección de malas hierbas en césped basada en imágenes) del año 2019 el más citado con 134.

Tabla 1
Ranking de patrocinadores financieros por número de documentos

Patrocinador financiero	Documento
National Natural Science Foundation of China	289
National Key Research and Development Program of China	72
National Science Foundation	59
National Institutes of Health	54
Horizon 2020 Framework Programme	46
National Research Foundation of Korea	43
Fundamental Research Funds for the Central Universities	40
European Commission	29
Engineering and Physical Sciences Research Council	25
European Research Council	23

Nota. La tabla muestra la lista de patrocinadores financieros para los trabajos de investigación en el entorno RNC. Fuente: Adaptado de Web Scopus año 2023.

Tabla 2
Ranking de autores principales

Nombre de autor principal	Documentos
Schumann, A.W.	9
Muhammad, K.	8
Baik, S.W.	6
Boyd, N.S.	6
Sharpe, S.M.	6
Wei, S.	6
Ampatzidis, Y.	5
Domínguez-Rodrigo, M.	5
Linares-Barranco, A.	5
Mohana	5

Nota. La tabla muestra la lista de investigadores que generan una mayor cantidad de literatura publicada. Fuente: Adaptado de Web Scopus año 2023.

En el caso de la investigación bibliométrica de Prahani et al, el autor Schumann, A.W. no figura en el ranking, siendo el que ocupa esta posición el autor He. En la investigación de Cheng y Deng tampoco aparece el autor considerado en la presente investigación, siendo los autores chinos Zhang, Y y Sun, X los ocupantes de esta posición.

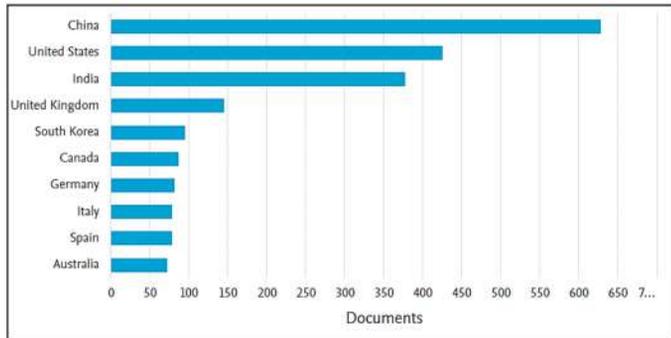
Análisis de documentos por país o territorio

Según la figura 11, China predomina como el país que más documentos ha publicado, siendo un total de 628, coincidiendo con las investigaciones de Prahani et al. y Chen y Deng, donde también figura el país de China como el principal aportante de documentos investigativos, le sigue Estados Unidos de América con 425 y a

continuación la India con 377. Obsérvese además que el único país de idioma español que figura en este top 10 es precisamente España, ubicado en la posición número nueve con 78 documentos publicados.

Figura 11

Cantidad de documentos publicados por país o territorio



Nota. La figura muestra la cantidad de documentos publicados por país o territorio. Fuente: Web Scopus año 2023.

En este punto, se debe mencionar a nuestro país Perú que figura con un aporte de 13 investigaciones, siendo la investigación titulada “Imagen del tamaño de la herida: Listo para la evaluación y el seguimiento inteligente”, desarrollado por investigadores de la Pontificia Universidad católica del Perú en colaboración con investigadores de empresas extranjeras, la que muestra el mayor número de citas, 23 en total desde su publicación en Scopus en el año 2021.

Tendencias futuras de las RNC en visión artificial

En primera instancia, resulta necesario continuar mejorando las sólidas arquitecturas existentes, incorporando lecciones aprendidas de modelos previos. Debería profundizarse en la exploración de técnicas de aprendizaje por transferencia y preentrenamiento para aprovechar conocimientos de conjuntos de datos extensos.

Para aumentar la robustez y la capacidad de generalización de los modelos RNC, se observa la tendencia a la integración de técnicas como el entrenamiento adversarial, la adaptación de dominios y el aprendizaje autosupervisado. La creación y mantenimiento de conjuntos de datos de reconocimiento facial diversos y completos resultan vitales para mejorar el entrenamiento y la evaluación de los modelos en situaciones del mundo real, tal como señala (Barcic et al., 2023).

La atención debe centrarse en el diseño eficiente de modelos, así como en técnicas de compresión y aceleración de hardware para optimizar las arquitecturas de las RNC en tiempo real y en entornos con limitados recursos. La adaptación de modelos y aprendizaje en base a nueva data será posible si se pone mayor atención a enfoques de aprendizaje continuo.

Actualmente se viene produciendo de manera creciente lanzamientos de marcos de software para CPU y GPU de alto consumo energético, incrementando los

niveles de contaminación ambiental, ante ello Venieris Stylianos propuso que el hardware basado en FPGA (Field- Programmable Gate Array) podría ser una plataforma alternativa para integrarse en los sistemas de aprendizaje profundo, ofreciendo un equilibrio ajustable entre rendimiento y consumo de energía, conforme lo indican en su artículo Chen y Deng (2020). Así mismo, estos autores destacan la relación estrecha entre la detección de objetivos, el análisis de vídeo y la comprensión de imágenes, en estos tiempos por lo que mencionan la evolución de los métodos tradicionales de detección de objetos hacia enfoques más potentes y avanzados, como el aprendizaje profundo.

Hoy en día el accionar de esta tecnología RNC se ha extendido a muchos campos donde se trabaje con visión artificial: robótica, medicina, minería, etc., siendo por ello necesario poner mayor atención a los logros y evoluciones en este campo.

Conclusiones

Las redes neuronales convolucionales (RNC) son un tipo específico de arquitectura de redes neuronales profundas diseñadas especialmente para el procesamiento de datos de tipo gráfico, como imágenes. En el contexto de la visión artificial, las RNCs han demostrado ser particularmente efectivas y se han convertido en la elección predominante para tareas de reconocimiento de imágenes, clasificación de objetos, detección de patrones y otras aplicaciones relacionadas con el procesamiento de imágenes. El análisis de la aplicación de las RNC en el campo de la visión artificial fue expuesto desde una perspectiva bibliométrica, examinándose 5 características, mediante la base de datos Scopus, para 5 años (2018 – 2022).

En la primera característica analizada se destaca un significativo aumento en la producción de artículos sobre RNC en el área de la visión artificial que refleja una tendencia consistente sugiriendo que la investigación en RNC posiblemente esté alcanzando la etapa de madurez. En un contexto de auge en el aprendizaje profundo, se destaca la urgencia de comprender y aplicar las RNC en diversos campos, con la necesidad de aumentar la experimentación para innovar en este tipo de red neuronal.

La segunda característica, revela que los campos principales de estudio en la literatura sobre RNC abarcan diversas áreas, como ciencias de la computación, ingeniería, matemáticas, medicina, física y agricultura, siendo que la ciencia de la computación destaca como la principal área de investigación seguida por ingeniería. El documento más citado en ciencias de la computación pertenece a Voulodimos et al. (2018), destaca por su revisión de esquemas de aprendizaje profundo aplicados en desafíos de visión por computadora.

En cuanto a la tercera característica, el análisis de instituciones de investigación con filiación de autores destaca la prominencia de la Academia China de Ciencias como líder en la producción de artículos sobre RNC. Este liderazgo se refleja en la colaboración y la diversidad de

áreas de investigación. La presencia significativa de instituciones chinas entre las diez principales sugiere un avance notable en la investigación sobre RNC en China en este quinquenio. Además, se indica que la National Natural Science Foundation of China es el principal patrocinador financiero de este tipo de investigaciones.

Para la cuarta característica, la determinación de los autores principales indica que Schumann, A.W. encabeza la lista con 9 publicaciones, destacando su artículo sobre aprendizaje profundo para la detección de malas hierbas en césped. Sin embargo, en investigaciones bibliométricas previas, las clasificaciones de autores principales difieren, destacando la importancia de considerar múltiples estudios para obtener una visión completa del panorama investigativo en un campo específico.

La última característica, se destaca a China como el líder indiscutible en la publicación de documentos sobre RNC, Estados Unidos sigue en segundo lugar seguido por India. España es el único país de habla hispana en el top 10, ocupando la novena posición. Perú figura en la lista con 13 investigaciones sobre el tema.

Finalmente, decir que este análisis sienta las bases para seguir investigando y desarrollando arquitecturas o modelos de RNC más precisos y robustos en el área de la visión artificial.

Fuente de financiamiento

Los fondos para la realización de esta investigación fueron proporcionados por la Universidad Nacional Daniel Alcides Carrión (UNDAC) Pasco – Perú, como parte de su compromiso con el fomento y apoyo a la investigación científica y el desarrollo académico.

Contribución de los autores

M. A. D. C. R.: se encargó de la redacción del análisis e interpretación de datos, así como de la elaboración de la parte teórica del informe final y del artículo científico.

Williams Muñoz Robles: se dedicó a la elaboración de tablas y figuras estadísticas, participando además en la redacción de los informes y del artículo científico.

M. A. T. M.: se encargó de la sistematización y redacción del marco teórico, así como de la redacción de la contrastación de los resultados con las bases teóricas.

Conflicto de Interés

Los autores de este artículo declaran no tener ningún conflicto de intereses relacionado con la investigación, la redacción o la publicación del mismo.

Referencias Bibliográficas

Aloysius, N., & Geetha, M. (2017). A review on deep convolutional neural networks. *2017 International Conference on Communication and Signal Processing (ICCSP)*, 0588-0592.

<https://doi.org/10.1109/ICCSP.2017.8286426>

Barcic, E., Grd, P., & Tomicic, I. (2023). *Convolutional Neural Networks for Face Recognition: A Systematic Literature Review*. <https://doi.org/10.21203/rs.3.rs-3145839/v1>

Borrella, B. (2022). *Introducción a la visión artificial: Procesos y aplicaciones* [Tesis de Ingeniería Matemática, Universidad Complutense]. <https://hdl.handle.net/20.500.14352/3290>

Chen, H., & Deng, Z. (2020). Bibliometric Analysis of the Application of Convolutional Neural Network in Computer Vision. *IEEE Access*, 8, 155417-155428. <https://doi.org/10.1109/ACCESS.2020.3019336>

Datascientest. (2022, marzo 7). Perceptrón: ¿qué es y para qué sirve? *Formación en ciencia de datos* | DataScientest.com. <https://datascientest.com/es/perceptron-que-es-y-para-que-sirve>

Durán, J. (2017). *Redes neuronales convolucionales en R: Reconocimiento de caracteres escritos a mano* [Tesis de Ingeniería Robótica, Universidad de Sevilla]. <https://idus.us.es/handle/11441/69564>

Forsyth, D., & Ponce, J. (2002). *Computer vision. A modern approach*. Pearson Editorial. http://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Computer%20Vision%20A%20Modern%20Approach%20-%20Forsyth%20,%20Ponce.pdf

García, E. (2019). *Introducción a las redes neuronales de convolución. Aplicación a la visión por ordenador* [Tesis, Universidad de Zaragoza]. <https://zaguan.unizar.es/record/87398?ln=es>

García, I., & Caranqui, V. (2015). La visión artificial y los campos de aplicación. *Tierra Infinita*, 1(1), Article 1. <https://doi.org/10.32645/26028131.76>

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (Versión 1). arXiv. <https://doi.org/10.48550/arXiv.1512.03385>

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications* (Versión 1). arXiv. <https://doi.org/10.48550/arXiv.1704.04861>

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2016). *Densely Connected Convolutional Networks* (Versión 5). arXiv. <https://doi.org/10.48550/arXiv.1608.06993>

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105. https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>

Prahani, B., Imah, E., Maureen, I., Rakhmawati, L., & Saphira, H. (2023). Trend and Visualization of Artificial Intelligence Research in the Last 10 Years. *TEM Journal*, 918-927. <https://doi.org/10.18421/TEM122-38>

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications*, 9(1), 12.

- <https://doi.org/10.3390/publications9010012>
Ringa Tech (Director). (2021, agosto 16). *Redes Neuronales Convolucionales—Clasificación avanzada de imágenes con IA/ML (CNN)*.
<https://www.youtube.com/watch?v=4sWhhQwHqg>
- Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (Versión 6). arXiv.
<https://doi.org/10.48550/arXiv.1409.1556>
- Steemit. (2019, enero 11). *Breve Historia de las Redes Neuronales Artificiales* (Artículo 1). Steemit.
<https://steemit.com/spanish/@iars.geo/breve-historias-de-las-redes-neuronales-artificiales-articulo-1>
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). *Rethinking the Inception Architecture for Computer Vision* (Versión 3). arXiv.
<https://doi.org/10.48550/arXiv.1512.00567>
- Tan, M., & Le, Q. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* (arXiv:1905.11946). arXiv.
<https://doi.org/10.48550/arXiv.1905.11946>
- Tomás, V., & Tomás, V. (2018). La Bibliometría en la evaluación de la actividad científica. *Hospital a Domicilio*, 2(4), 145.
<https://doi.org/10.22585/hospdomic.v2i4.51>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018, 7068349.
<https://doi.org/10.1155/2018/7068349>
- Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., & Atkinson, P. (2022). UNetFormer: A UNet-like Transformer for Efficient Semantic Segmentation of Remote Sensing Urban Scene Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190, 196-214. <https://doi.org/10.1016/j.isprsjprs.2022.06.008>
- Xu, X., Tan, M., Corcoran, B., Wu, J., Boes, A., Nguyen, T., Chu, S., Little, B., Hicks, D., Morandotti, R., Mitchell, A., & Moss, D. (2021). 11 TeraFLOPs per second photonic convolutional accelerator for deep learning optical neural networks. *Nature*, 589(7840), 44-51.
<https://doi.org/10.1038/s41586-020-03063-0>
- Zhang, A., Lipton, Z., Li, M., & Smola, A. (2021). *Dive into Deep Learning* (Versión 5). arXiv.
<https://doi.org/10.48550/arXiv.2106.11342>