UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO

Facultad de Ingeniería Eléctrica

División de Estudios de Posgrado

DISEÑO DE ÍNDICES PARA LA RECUPERACIÓN DE OBJETOS MULTIMEDIA

TESIS

Que para obtener el grado de

DOCTOR EN CIENCIAS EN INGENIERÍA ELÉCTRICA

opción Sistemas Computacionales

Presenta

Bryan Eduardo Martínez Guzmán

Dr. José Antonio Camarena Ibarrola

Director de Tesis

Dr. Edgar Leonel Chávez González

Co-Director de Tesis

Morelia, Michoacán Enero 2025

Mis más sinceros agradecimientos a:

La División de estudios de posgrado de la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo, porque ha sido mi casa desde mis estudios de licenciatura, soy nicolaita de corazón.

Un agradecimiento especialmente a mi madre Jackeline Guzmán Reyes, y mi abuelita Victoria Reyes Valverde (†), por inculcarme los valores que me identifican, por su amor, confianza y apoyo incondicional con el cual he podido llegar a lograr mis metas. Quiero agradecer también a mi esposa Arely Guadalupe, por su cariño y confianza incondicional.

A mis asesores el Dr. José Antonio Camarena Ibarrola y el Dr. Edgar Leonel Chávez González por sus comentarios, recomendaciones y apoyo, los cuales me ayudaron en mis estudios de posgrado.

Al Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCyT), ahora la Secretaría de Ciencia, Humanidades, Tecnología e Innovación, por los apoyos financieros para cursar mis estudios de posgrado.

Y a muchas más personas valiosas que mi Dios me permitió conocer y trabajar desde el 2018.

¡¡¡Muchas a gracias a todos!!!

"We need to teach how doubt is not to be feared but welcomed. It's OK to say, 'I don't know'"

— Richard P. Feynman

Lista de Publicaciones

- Diseño de una función de pérdida perceptual basada en códigos de Hadamard (*Design of a brief perceptual loss function with Hadamard codes*) Quiroz, Bryan; Martínez, Bryan; Camarena-Ibarrola, Antonio; Chávez, Edgar. Multimedia Tools and Applications, 2024 Springer
- Recuperación de formas mediante coincidencia de polígonos (*Shape Retrieval through Polygon Matching*) Martínez, Bryan; Figueroa, Karina; Camarena-Ibarrola, Antonio. Conferencia Internacional Mexicana de Inteligencia Artificial (MICAI 2022)

Resumen

En la era del big data y la inteligencia artificial, el reconocimiento de imágenes se ha convertido en un desafío por la vasta cantidad de información multimedia generada de las últimas décadas. Manejar toda esta información es un reto para los métodos tradicionales de Recuperación de Imágenes Basada en Contenido (Content-Based Image Retrieval, CBIR por sus siglas del inglés), porque a menudo se ven limitados por problemas de escalabilidad, eficiencia computacional o decremento del recall en tareas de búsqueda y/o clasificación.

En esta tesis doctoral se presenta una metodología de investigación para la recuperación eficiente de imágenes en grandes bases de datos. Esta metodología supera las limitaciones de los sistemas CBIR tradicionales en términos de escalabilidad, eficiencia computacional y estabilidad del recall. La metodología propuesta se fundamenta en tres contribuciones clave:

- Una técnica avanzada de compresión de características profundas utilizando códigos de Hadamard, con la cual se logra una reducción del 75 % en el uso de memoria sin comprometer la calidad de la representación de las imágenes.

- Un novedoso esquema de indexación basado en matrices de Hadamard, con el cual se mejora la velocidad de búsqueda en un 20 % y proporciona robustez frente a transformaciones y ataques adversarios.

- Un algoritmo de optimización de redes neuronales convolucionales, con el cual se seleccionan eficientemente las neuronas más valiosas, y se reduce substancialmente el uso de memoria.

Los resultados experimentales demuestran la superioridad de la metodología propuesta frente al estado del arte, no solo por las mejoras en uso de memoria y velocidad,

sino también porque, con esta metodología, se logra mantener un recall constante y competitivo en diversos escenarios de evaluación. Estas contribuciones ayudan en el desarrollo de sistemas de Inteligencia Artificial más eficientes, robustos y adaptables a los desafíos del big data.

Palabras clave Recuperación de Imágenes Basada en Contenido, Redes Neuronales Convolucionales, Códigos de Hadamard, Índices, Navegación multimedia.

Abstract

In the era of big data and artificial intelligence, image recognition has become a challenge due to the vast amount of multimedia information generated in recent decades. Managing all this information is a challenge for traditional Content-Based Image Retrieval (CBIR) methods, because they are often limited by scalability issues, computational efficiency, or decreased recall in search and/or classification tasks. In this doctoral thesis, a research methodology for the efficient retrieval of images in large databases is presented.

This methodology overcomes the limitations of traditional CBIR systems in terms of scalability, computational efficiency, and recall stability. The proposed methodology is based on three key contributions:

• An advanced deep feature compression technique using Hadamard codes, which achieves a 75% reduction in memory usage without compromising image rendering quality.

• A novel indexing scheme based on Hadamard matrices, which improves search speed by 20% and provides robustness against transformations and adversary attacks.

• An algorithm for optimizing convolutional neural networks, with which the most valuable neurons are efficiently selected, and memory usage is substantially reduced.

The experimental results demonstrate the superiority of the proposed methodology over the state of the art, not only because of the improvements in memory and speed usage, but also because, with this methodology, it is possible to maintain a constant and competitive recall in various evaluation scenarios. These contributions help in the development of more efficient, robust, and adaptable Artificial Intelligence systems to the challenges of big data.

Contenido

Lista de Figuras

Lista de Tablas

Lista de Algoritmos

Lista de Símbolos

α	Índice de elemento en una base de datos
β	Vector de características
λ	n-ésima raíz unitaria
ϕ	Función de transformación
φ_ℓ	Función escalar compleja tipo ℓ
$\mathbb{C}$	Conjunto de números complejos
H	Matriz de Hadamard
W	Matriz de transformación lineal
N	Conjunto de neuronas
N_k	Subconjunto de k neuronas
O	Conjunto de objetos
q	Imagen de consulta
k	Número de vecinos más cercanos o número de neuronas
n	Número de clases o dimensión
w_i	Peso asignado a la neurona i
a_i	Número de activaciones de la neurona i
$S[p(x)]$	Entropía de Shannon

Lista de Acrónimos

AI	Artificial Intelligence (Inteligencia Artificial)
AP	Angular Pattern (Patrón Angular)
BAP	Binary Angular Pattern (Patrón Angular Binario)
CBIR	Content-Based Image Retrieval (Recuperación de Imágenes Basada en Contenido)
CNN	Convolutional Neural Network (Red Neuronal Convolucional)
CSS	Curvature Scale Space (Espacio de Escala de Curvatura)
DHF	Deep Hadamard Features (Características Profundas de Hadamard)
DSW	Dynamic Space Warping (Alineamiento Espacial Dinámico)
DTW	Dynamic Time Warping (Alineamiento Temporal Dinámico)
EDN	Equal Distance Normalization (Normalización de Distancias Iguales)
FAISS	Facebook AI Similarity Search (Búsqueda de similitudes de IA en Facebook)
GPU	Graphics Processing Unit (Unidad de Procesamiento Gráfico)
TPU	Tensorial Processing Unit (Unidad de Procesamiento Tensorial)
NPU	Neural Processing Unit (Unidad de Procesamiento Neuronal)
HNSW	Hierarchical Navigable Small World (Mundo Pequeño de Navegación Jerárquico)
HSP	Half-Space Proximal (Mitad Espacial Proximal)
ILSVRC	ImageNet Large Scale Visual Recognition Challenge (Desafío de Reconocimiento Visual a Gran Escala de ImageNet)
MSE	Mean Square Error (Error Cuadrático Medio)
OAN	Object Area Normalization (Normalización del Área del Objeto)
PCA	Principal Component Analysis (Análisis de Componentes Principales)
PSM	Partial Shape Matching (Coincidencia de Formas Parciales)
RAM	Random Access Memory (Memoria de Acceso Aleatorio)
ReLU	Rectified Linear Unit (Unidad Lineal Rectificada)
RFE	Recursive Feature Elimination (Eliminación de Características Recursiva)

SBS	Sequential Backward Selection (Selección Secuencial hacia Atrás)
SCN	Shape Classification Network (Red de Clasificación de Formas)
TAR	Triangle-Area Representation (Representación de Área Triangular)
TF-IDF	Term Frequency-Inverse Document Frequency (Término de Frecuencia-Frecuencia de documento inversa)
TPU	Tensor Processing Unit (Unidad de Procesamiento Tensorial)
ZMD	Zernike Moment Descriptor (Descriptor de Momentos de Zernike)

Capítulo 1

Introducción

"La imaginación es más importante que el conocimiento. El conocimiento es limitado, mientras que la imaginación abarca el mundo entero, estimulando el progreso, dando nacimiento a la evolución."

Albert Einstein (1879-1955)
Físico teórico y científico

En las últimas décadas, el área de visión computacional ha experimentado avances significativos gracias al desarrollo de hardware especializado como las Unidades de Procesamiento Gráfico (GPUs), Unidades de Procesamiento Tensorial (TPUs), Unidades de Procesamiento Neuronal (NPUs), así como mejoras a los sistemas de inteligencia artificial, y la disponibilidad de grandes conjuntos de datos. Estas herramientas han permitido a los investigadores de la actualidad, atacar problemas complejos que anteriormente se consideraban irresolubles. Esto ha dado lugar a aplicaciones capaces de procesar imágenes, audio y video, de manera simultánea, eficiente y veloz.

1.1. Estado del Arte

La Recuperación de Imágenes Basada en Contenido (Content-Based Image Retrieval, CBIR) es un área fundamental de visión computacional que ha sentado las bases a numerosas aplicaciones modernas, como lo son, sistemas de reconocimiento, sistemas de vigilancia automática y sistemas de anotación automática entre otros [Smeulders00, Datta08, Liu07].

Con el fin de mostrar cómo esta área de interés ha adquirido una gran importancia en la inteligencia artificial y el big data a través del tiempo, autores como Zhao et al. [Zhao24], Wan et al. [Wan14], Mikolajczyk et. al [Mikolajczyk05], Zhang et al. [Zhang04] y Lowe et al. [Lowe04] han elaborado diversos compendios con los trabajos más relevantes desde la década de los 1990. A partir de esta década, las técnicas de recuperación han evolucionado desde arquitecturas simples hasta arquitecturas complejas, con las cuales es posible manejar volúmenes de datos cada vez mayores, y de una mejor manera.

Las técnicas de recuperación mencionadas en este trabajo de investigación emplean las propiedades intrínsecas de los objetos como color, textura, tamaño, curvatura de la forma, y también descriptores de Fourier [Gonzalez02, Zhang04] como los principales componentes para recuperar imágenes similares de bases de datos grandes. Porque estas propiedades son comúnmente utilizadas para este fin desde hace un par de décadas, tal y como se menciona con Wan, Mikolajczyk y Lowe entre otros autores.

1.1.1. Técnicas fundamentales

El color y la textura poseen una riqueza informativa muy alta, esta es la principal razón por la cual estas propiedades son muy utilizados en diversas aplicaciones [Swain91, Manjunath96]. Autores como Rui et al. [Rui99] en sus experimentos han combinado varias propiedades de los objetos, entre ellas el color y la textura, de esta manera han podido alcanzar resultados muy impresionantes que han marcado tendencia en CBIR.

La curvatura de la forma ha ganado relevancia por su complejidad de procesamiento y su relación con la percepción humana [Zhang04, Belongie02]. Los seres humanos tendemos a relacionar objetos por su forma principalmente, después por su color o textura [Biederman87]. Por ejemplo, las manzanas se pueden reconocer más fácilmente por su forma

característica, que por su color, a pesar de que existen una variedad de manzanas rojas, verdes, amarillas, negras (del Tíbet) o blancas (de Galicia) [Janick19].

La literatura sobre clasificación de objetos mediante la forma de su curvatura es extensa, como lo muestran Rui et. al [Rui99] y Belongie et. al [Belongie02]. En esta literatura es común utilizar el algoritmo Dynamic Time Warping (DTW) [Kumar21] para comparar formas; sin embargo, varios autores, en especial Rui y Belongie descubrieron que en conjuntos de datos grandes, el algoritmo DTW eleva el tiempo de procesamiento exponencialmente, lo cual es un problema para bases de datos masivas. Un par de años adelante DTW fue reemplazado por el algoritmo Dynamic Space Warping (DSW) [Alajlan11], con la esperanza de reducir el tiempo de procesamiento, sin embargo, esto no fue posible porque DSW solo es ligeramente mejor que DTW, por lo tanto, falla con datos masivos.

En los últimos años se han incorporado más descriptores al estado del arte, entre ellos han sobresalido los momentos de Zernike y la transformada de Fourier. Estos descriptores han demostrado ser adaptables y eficientes en la recuperación de objetos de grandes bases de datos [Bartolini05], porque generan puntos clave directamente de la curvatura de la forma. Esta idea es sorprendente, porque al discernir los objetos por su forma y tomar muestras equidistantes, disminuye radicalmente la cantidad de operaciones y con ello reduce el tiempo de procesamiento [Zhang03, Alajlan07].

1.1.2. Avances recientes

Por otro lado, los sistemas de inteligencia artificial tuvieron un importante avance con la introducción de las Redes Neuronales, especialmente con las Redes Neuronales Convolucionales (Convolutional neural network, CNNs por sus siglas de inglés). El trabajo hecho por Yann LeCun et. al [LeCun10] pionero de este campo, han permitido que surjan nuevos sistemas de reconocimiento de objetos bastante eficientes y rápidos, como es el caso de YOLO (You Only Look Once) [Redmon16], y otros como:

- RetinaNet (2018) por Tsung-Yi et. al. [Lin17]. RetinaNet al igual que YOLO se basa en una estructura de CNN para reconocer objetos de diversos tamaños.

- Google Spinet (2019) por Xianzhi et. al. [Du20]. Google Spinet propone una arquitec-

tura novedosa llamada “Scale-Permuted” en la que se alternan diversos tamaños de capas convolucionales dentro de su aquitectura.

- Facebook DETR (2020) por Nicolas et. al. [Carion20]. Facebook DETR propone el uso de transformadores para detectar imágenes.

- CLIP por Radford et. al. [Radford21]. CLIP es una red neuronal que aprende conceptos visuales del lenguaje natural.

- Socratic Models (2022) por Zeng et. al. [Zeng22]. Socratic Models es un framework que se compone por múltiples modelos previamente entrenados para realizar tareas multimodales.

Estos sistemas son precisos, eficientes y veloces, sin embargo, aún enfrentan desafíos en su robustez frente a ataques adversarios debido (en gran parte) al clasificador One-Hot en su arquitectura. One-Hot es muy utilizados en CNNs porque permite identificar con un solo bit la clase a la que pertenece la imagen de consulta, desafortunadamente, One-Hot puede sufrir alteraciones con el mínimo esfuerzo [Goodfellow14].

Mohammed et al. [Mohammed23] preocupados por estos inconvenientes, han realizado un análisis exhaustivo de las vulnerabilidades y desafíos de seguridad en sistemas de recuperación de imágenes basados en aprendizaje profundo, en su trabajo destacan la importancia de desarrollar soluciones robustas contra ataques adversarios y otras amenazas potenciales.

Finalmente, además de la eficiencia y eficacia, los sistemas CBIR necesitan ser veloces, esta cualidad la han ganado aquellos sistemas que implementan técnicas sofisticadas de búsqueda y recuperación, así como la incorporación de estructuras de datos escalables como árboles-M, árboles-kd, tablas Hash, e indexadores bien establecidos como Faiss de Facebook (2016) propuesto por Matthijs et. al. [Douze24] y HNSWLIB (2018) propuesto por Yu et. al. [Malkov18].

Derivado de la información presentada en esta sección, se puede afirmar que los sistemas CBIR actuales están conformados por la unión de todos estos componentes, y la calidad de su respuesta, depende de que tan eficientes y robustos sean sus elementos.

1.2. Definición del Problema

En la era actual del big data, las plataformas sociales, los sistemas de vigilancia, las aplicaciones médicas y los sistemas de servicios de comercio electrónico entre otros, generan y almacenan diariamente millones de imágenes, este incremento de información ha creado una necesidad urgente de sistemas de recuperación de imágenes eficientes y precisos, capaces de procesar y analizar grandes volúmenes de datos en poco tiempo.

La recuperación de imágenes basada en contenido en entornos de big data representa un desafío de la visión computacional, el aprendizaje automático y la gestión de datos masivos. El problema a resolver en esta investigación doctoral se centra en la recuperación eficiente de las imágenes más similares a una imagen de consulta, dentro de una base de datos de gran escala.

1.2.1. Formulación Matemática del Problema

Sea el problema a resolver formalmente definido como:

- Una base de datos de imágenes $D = \{x_i\}$ donde $i = 1, \ldots, n$, donde:
  - Cada x_i representa una imagen individual
  - n denota el número total de imágenes en la base de datos
- Una imagen de consulta q
- Un número k que especifica el número de las k-imágenes más similares a recuperar

El objetivo es desarrollar una metodología que optimice la función:

$$K(q, D, k) = \{x_{i=1}, x_{i=2}, ..., x_k\} \subset D \tag{1.1}$$

Donde K representa el conjunto de las k imágenes más similares a q de la base de datos D, según la métrica de similitud definida (ver Figura 1.1).

Figura 1.1: Ejemplo del problema a resolver

En la Figura 1.1 se muestra un ejemplo a del problema a resolver en este trabajo de investigación. La sección a) representa a una imagen de consulta q arbitraria, en este caso corresponde a un pez naranja que el usuario desea buscar. Como se puede apreciar, esa figura posee características distinguibles como son el color naranja, la forma y los patrones particulares que hacen al pez que sea distintivo de otros objetos. En la sección b) se muestra la base de datos de la cual se va a hacer la recuperación de imágenes que más se le parezcan a la consulta.

Durante la búsqueda de las imágenes semejantes al pez naranja de consulta, la función K identificará otras imágenes que compartan características similares, como lo es el color, la forma del pez, y patrones específicos correspondientes a peces, independientemente de variaciones en iluminación, ángulos, rotaciones, escalamientos, poses o cualquier otra perturbación. Por este motivo, imágenes que tengan peces naranja, incluso objetos naranjas, tendrán alta similitud, a diferencia de otros objetos como edificios grises o vehículos azules, los cuales tendrán baja similitud.

Finalmente, en la sección c) de esta Figura 1.1, se muestra el resultado final con las k-imágenes más semejantes, ordenadas por su grado de similitud con respecto el pez naranja de consulta. Como se puede apreciar en este ejemplo, estos son los pasos que realizan los sistemas CBIR para buscar imágenes semejantes, y que pueden ser aplicados a otros problemas como es el caso de la navegación web. En este caso se realiza el mismo procedimiento de búsqueda, la diferencia consiste en la cantidad de información que se maneja.

1.2.2. Desafíos

El problema de la recuperación de imágenes basada en contenido en la era del big data, presenta una serie de desafíos profundamente interconectados que requieren una solución integral y sistemática. Consideremos estos desafíos usando un caso práctico, imagine una plataforma de comercio electrónico que maneja millones de imágenes de productos. Cuando un usuario sube una foto de un producto que desea encontrar, los sistemas CBIR se enfrentan a problemas de:

1. **Escalabilidad**: En el contexto actual, donde las bases de datos contienen millones o incluso miles de millones de imágenes, los métodos tradicionales de búsqueda se vuelven computacionalmente costosos. Este escenario demanda el desarrollo e implementación de técnicas de indexación y búsqueda innovadoras, que puedan escalar eficientemente con el crecimiento continuo de las bases de datos.

2. **Eficiencia computacional**: Las aplicaciones del mundo real exigen que la recuperación de objetos similares se realice con una rapidez excepcional. Este requisito necesita no solo de algoritmos altamente optimizados, sino también de estructuras de datos robustas que soporten operaciones de búsqueda eficientes.

3. **Representación de características**: La efectividad de los sistemas CBIR dependen en su manera de representar las imágenes. Estas representaciones deben lograr un balance óptimo entre la captura eficaz del contexto y la reducción del tamaño de los datos, manteniendo la información semántica esencial, es decir, se deben procesar las imágenes a modo que sean lo más compactas posible y que sean altamente informativas.

4. **Robustez**: Un desafío particularmente complejo emerge de la necesidad de mantener la precisión del reconocimiento bajo diversas condiciones adversas. Esto incluye variaciones en la iluminación, transformaciones afines como escalamiento y rotación, deformaciones, y la creciente amenaza de ataques adversarios. La capacidad de mantener un rendimiento consistente bajo estas condiciones es importante para la aplicabilidad de los sistemas CBIR robustos.

5. **Uso eficiente de la memoria**: La gestión óptima de recursos computacionales requiere el desarrollo de técnicas y métodos que minimicen el uso de memoria en dos aspectos críticos: el almacenamiento de las representaciones de imágenes y el proceso dinámico de recuperación. Este desafío se magnifica con el crecimiento de las bases de datos y la necesidad de mantener tiempos de respuesta reducidos.

Estos desafíos no existen de manera aislada, sino que se interrelacionan y afectan mutuamente, por ejemplo, al obtener representaciones más compactas, se reduce memoria, lo cual impacta directamente en la eficiencia y la escalabilidad del sistema, mientras que la necesidad de robustez puede afectar la eficiencia computacional. Esta interdependencia subraya la importancia de abordar estos desafíos de manera holística, porque cada solución particular afecta al rendimiento global de la metodología propuesta.

La complejidad de estos desafíos, lejos de ser un obstáculo insuperable, representa una oportunidad para el desarrollo de soluciones innovadoras que puedan mejorar significativamente el estado actual de los sistemas CBIR. El abordar estos desafíos de manera efectiva no solo se avanzará el campo en cuestión, sino que también abrirá nuevas posibilidades para aplicaciones prácticas en diversos dominios.

Por lo tanto, el problema planteado en este tema de investigación trasciende la recuperación de imágenes semejantes, porque representa un desafío multidisciplinar que requiere la integración de conceptos de visión computacional, aprendizaje automático y estructuras de datos, entre otras áreas y disciplinas relacionadas.

La solución propuesta en este trabajo de investigación hace frente a estos desafíos con la incorporación de técnicas y métodos escalables, robustos, con los cuales se puede gestionar de manera eficiente el uso de recursos computacionales. Esto con el fin de asegurar que una imagen sea correctamente reconocida por sus propiedades, y en el supuesto de que exista algún tipo de ruido, las propiedades de las imágenes se deberían de sobreponer. Un ejemplo que engloba estos desafíos es el caso de la Figura 1.2. Esta figura corresponde a un perro labrador porque, todas las características del labrador son claramente distinguibles: como su pelaje brillante, sus orejas caídas (características de la raza), y su expresión facial amigable. Un sistema CBIR vulnerable no es capaz de reconocer correctamente al labrador,

si la imagen original fue expuesta a perturbación como se muestra en la sección b), esta acción provoca una mala clasificación y se le asigna erróneamente la etiqueta “gato” como se muestra en la sección c). En este trabajo de investigación se implementan tecnologías robustas para aminorar este efecto.

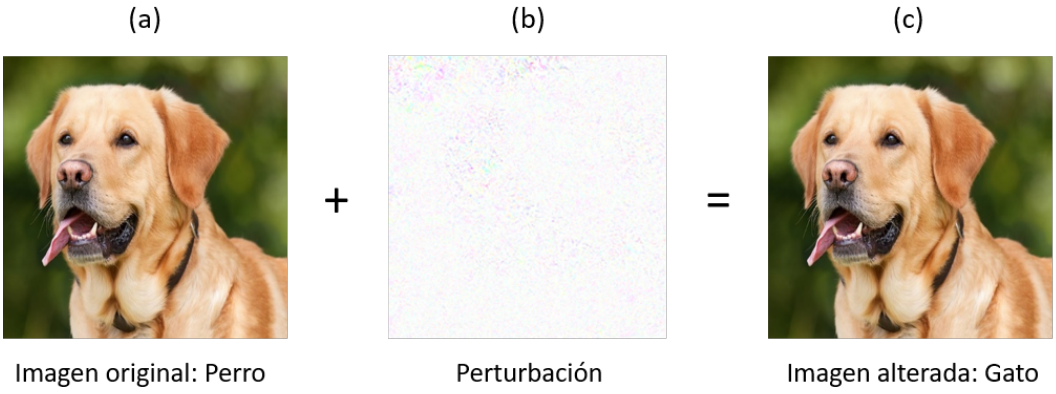


Figura 1.2: Ejemplo de los desafíos que intervienen en los sistemas CBIR

1.3. Justificación

Los sistemas de Recuperación de Imágenes Basada en Contenido (CBIR) enfrentan desafíos fundamentales en la era del big data. Estos sistemas deben gestionar eficientemente repositorios que contienen miles de millones de imágenes, incluso billones de imágenes, también deben procesar múltiples consultas sin conocimiento previo del contenido y realizar cálculos complejos de similitud para seleccionar los vecinos más cercanos. La magnitud de estos retos se amplifica con el crecimiento exponencial de datos en aplicaciones modernas, que van desde motores de búsqueda hasta sistemas de diagnóstico médico y más.

1.3.1. Limitaciones de los Enfoques Actuales

Las soluciones actuales, aunque son prometedoras presentan limitaciones significativas que requieren atención inmediata. El uso de Redes Neuronales Convolucionales (CNNs) para la extracción de características profundas, si bien ha demostrado ser efectivo

en la comparación de objetos, conlleva una alta dimensionalidad. Esta dimensionalidad resulta en un excesivo uso de memoria y capacidad de procesamiento, generando tiempos de respuesta que pueden resultar no apropiados para aplicaciones en tiempo real como señalan Wang et al. [Wang23b] en su análisis de eficiencia computacional. Wang señala que estos factores limitan severamente la escalabilidad de los sistemas CBIR actuales.

Una limitación importante se relaciona con la sensibilidad de dichos sistemas frente a variaciones en los datos de entrada y sensibilidad ante ataques adversarios. Los estudios realizados por Quiroz et al. [Quiroz24] y Dong et al. [Dong22] sugieren que las representaciones aprendidas por modelos convolucionales pueden verse afectadas por cambios sutiles en los datos, lo cual impacta directamente en la precisión. Existen plataformas en línea dedicadas a la recuperación de imágenes semejantes, que presentan estas limitaciones e incluso más, como es el caso de:

- **Google Images** `images.google.com`. Es el líder indiscutible en búsqueda de imágenes por varias. Sin embargo, su principal desafío radica en el equilibrio entre la privacidad del usuario y la efectividad de la búsqueda.
- **TinEye** `tineye.com`. Se destaca por su capacidad única para encontrar imágenes exactas y variaciones de una imagen específica; sin embargo, su principal inconveniente es el tiempo de cómputo.
- **Lenso.AI** `lenso.ai`. Representa la nueva generación de buscadores que utilizando inteligencia artificial avanzada. Su importancia se encuentra en su capacidad para entender contextos complejos y realizar búsquedas por contexto; sin embargo, sus principales inconvenientes es que no es tan robusto, ni tan preciso.
- **Getty Images** `gettyimages.com` y **Clarifai** `clarifai.com`. Estos son buscadores veloces, pero no son precisos, ni robustos.

En la Tabla 1.1 se muestran los resultados de la imagen de consulta 1.1 en estas cuatro plataformas: Google Images, TinEye, Lenso.AI y Getty Images. Para cada plataforma, se realizaron dos consultas: una versión sin alteraciones y otra con una ligera perturbación. Los resultados revelan variaciones sutiles en la capacidad de cada sistema para identificar

similitudes entre las imágenes. Mientras que algunas plataformas mostraron robustez frente a las perturbaciones, manteniendo resultados consistentes entre ambas versiones, otras exhibieron diferencias notables en los resultados de búsqueda.

	Imagen sin perturbación	**Imagen con perturbación**
a) *Google Images*		
b) *TinEye*		
c) *Lenso.AI*		
d) *Getty Images*		

Tabla 1.1: Recuperación de imágenes usando plataformas CBIR en línea

Las variaciones de la Tabla 1.1 evidencian las limitaciones de los sistemas actuales y resaltan la importancia de la metodología propuesta en esta tesis, en la cual se ofrece mayor robustez frente a perturbaciones gracias a la combinación de técnicas implementadas,

permitiendo mantener un recall consistente incluso en presencia de alteraciones en la imagen de consulta.

1.3.2. Oportunidades de Mejora

Los avances recientes documentados por Zhao et al. [Zhao24] en optimización de redes neuronales sugieren un amplio margen de mejora en diversos aspectos del campo. La eficiencia computacional de los sistemas CBIR puede incrementarse significativamente mediante nuevas arquitecturas optimizadas, mientras que la precisión en la recuperación de imágenes puede mejorarse a través de técnicas más sofisticadas de procesamiento. Además, la robustez frente a variaciones y perturbaciones puede fortalecerse mediante enfoques innovadores de entrenamiento y validación.

1.3.3. Hipótesis

Frente a estos desafíos y oportunidades, en esta tesis doctoral se propone una solución innovadora basada en la siguiente hipótesis:

> "Mediante la transferencia de conocimiento de redes neuronales convolucionales (CNN) entrenadas con grandes bases de datos, es posible generar representaciones comprimidas que, al utilizarse como índices de objetos, permiten realizar búsqueda y recuperación de imágenes similares en bases de datos desconocidas."

Esta hipótesis encuentra su fundamento en los avances recientes acerca del aprendizaje por transferencia y las técnicas de compresión de modelos documentadas por Li et al. [Li22b]. La optimización de arquitecturas neuronales para recursos limitados representa un camino prometedor para superar las limitaciones actuales de los sistemas CBIR.

La relevancia y pertinencia de esta investigación se sustenta en el aporte de este trabajo para abordar múltiples aspectos críticos del campo. En términos de optimización de recursos, se busca una reducción significativa de los requisitos de memoria y una disminución sustancial de la carga computacional, permitiendo un mejor aprovechamiento de recursos en sistemas CBIR reales.

La mejora en eficiencia constituye otro pilar fundamental de este trabajo, se prevé alcanzar una velocidad superior en la recuperación de imágenes con respecto a otros traba-

jos, acompañada de una escalabilidad mejorada para repositorios masivos. Esto permitiría obtener respuestas en más consistentes al instante, este es un aspecto muy importante para aplicaciones prácticas que se deriven de este trabajo.

El impacto potencial de esta investigación se extiende a múltiples dominios de aplicación, porque los motores de búsqueda podrían beneficiarse de sistemas más precisos y eficientes, con conocimiento previo, al igual que los sistemas de recomendación basados en contenido, podrían ofrecer sugerencias más relevantes y personalizadas. Además, la posibilidad de implementar aplicaciones sofisticadas en dispositivos con recursos limitados abre nuevas oportunidades para el comercio electrónico y el mundo del entretenimiento.

1.4. Motivación

Los seres humanos en la actualidad vivimos rodeados de un gran volumen de imágenes. Las redes sociales, la fotografía digital y las imágenes satelitales, entre otras fuentes, generan constantemente un flujo masivo de información que documenta y mapea nuestro mundo. Por este motivo es que las imágenes se han transformado en un lenguaje universal y una herramienta fundamental en nuestra vida cotidiana, porque influyen en la manera en que nos comunicamos, aprendemos, jugamos y entendemos nuestro entorno, algunos sistemas que están inmersos en nuestras actividades cotidianas son:

- Los sistemas de seguridad cuando deben identificar rápidamente a una persona, animal u objeto.
- Las plataformas de comercio electrónico, al momento de mostrar productos semejantes a lo que buscamos.
- O en las redes sociales, cuando los usuarios buscan contenido similar a una foto. En este caso, plataformas como Instagram o Facebook u otras, procesan miles de millones de fotos diariamente para atender a esta necesidad.

Cada uno de estos escenarios representa un desafío para los sistemas CBIR por el procesamiento de grandes volúmenes de datos, entre otros problemas críticos como lo son:

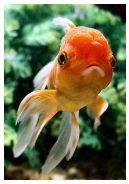
- **Velocidad vs Precisión**: Los sistemas de recuperación que son rápidos tienden usualmente a ser imprecisos, mientras que los precisos, son demasiado lentos para aplicaciones en tiempo real.
- **Uso de Memoria**: Las técnicas actuales requieren cantidades masivas de memoria, haciendo inviable su implementación en dispositivos con recursos limitados.
- **Escalabilidad**: El rendimiento se degrada significativamente al aumentar el tamaño de la base de datos.

Esta investigación surge para aminorar el efecto de estas limitaciones, y aprovechar todo el potencial que brindan los sistemas CBIR. La solución propuesta en este trabajo de investigación, es una metodología innovadora para; reducir el uso de memoria en un 75 %, mejorar la velocidad de búsqueda en un 20 % y mantener un alto rendimiento en diferentes situaciones de recuperación. Con estos aportes es posible recuperar imágenes basadas en contenido en el big data de una manera eficaz, eficiente, y rápida.

1.4.1. Ejemplos de Aplicación

Las aplicaciones que se pueden implementar y/o mejorar de las actuales, usando las bases sentadas en este trabajo de investigación, son muchas, por ejemplo:

- En el ámbito científico, considere el caso de un biólogo marino que fotografía un pez desconocido durante una inmersión. Al consultar la imagen obtenida, este no solo identifica la especie del pescado, sino que obtiene información adicional, como un análisis completo, incluyendo información taxonómica, hábitat natural y estado de conservación, también obtiene publicaciones científicas relevantes y bases de datos especializadas, entre más información importante para su área de estudio, y todo gracias al reconocimiento de una imagen, como se muestra a continuación:.

Consulta		Imágenes similares				
	Pescado					

- En las redes sociales, imagine a una persona buscando fotos de una fiesta usando como referencia un vaso tequilero distintivo. Al consultar esa imagen se podría traer información de las empresas que los han utilizado, eventos relacionados, personas y diferentes momentos de celebración. En este ejemplo, no solo se recuperaron imágenes donde aparece un vaso, sino que fueron establecidas varias conexiones con otros momentos memorables y otras personas, gracias al reconocimiento de una imagen, como se muestra a continuación:

- En el ámbito del comercio electrónico, imagine un usuario cualquiera buscando un par de tenis en línea, y este obtiene información relacionada con el modelo, color, medidas y otras variantes en diferentes tiendas, además de otros relacionados con la solicitud de tenis. Con esta información recuperada, el usuario podría tener acceso a una experiencia de compra personalizada, como se muestra a continuación:

En estos ejemplos se puede apreciar el potencial de esta investigación, no solo por la recuperación de las imágenes, también por la cantidad de información que se puede obtener al momento de consultar una imagen.

1.5. Objetivos de la Tesis

El objetivo principal de esta tesis es desarrollar una metodología robusta e innovadora para abordar el problema de la recuperación de imágenes basadas en contenido, con la cual se reduzca significativamente el uso de memoria y sea escalable a repositorios de imágenes muy grandes.

Los objetivos particulares son:

1. **Optimizar la memoria**. Reduciendo la cantidad de bytes que ocupan las características profundas para permitir el manejo eficiente de grandes bases de datos en memoria principal.

2. **Aumentar la velocidad**. Proponiendo e implementando técnicas robustas que permitan la indexación eficiente de imágenes, además de mantener recall comparable a métodos existentes

3. **Mejorar la escalabilidad**. Diseñando e implementando un indexador que particione el espacio de búsqueda de manera eficaz y eficiente.

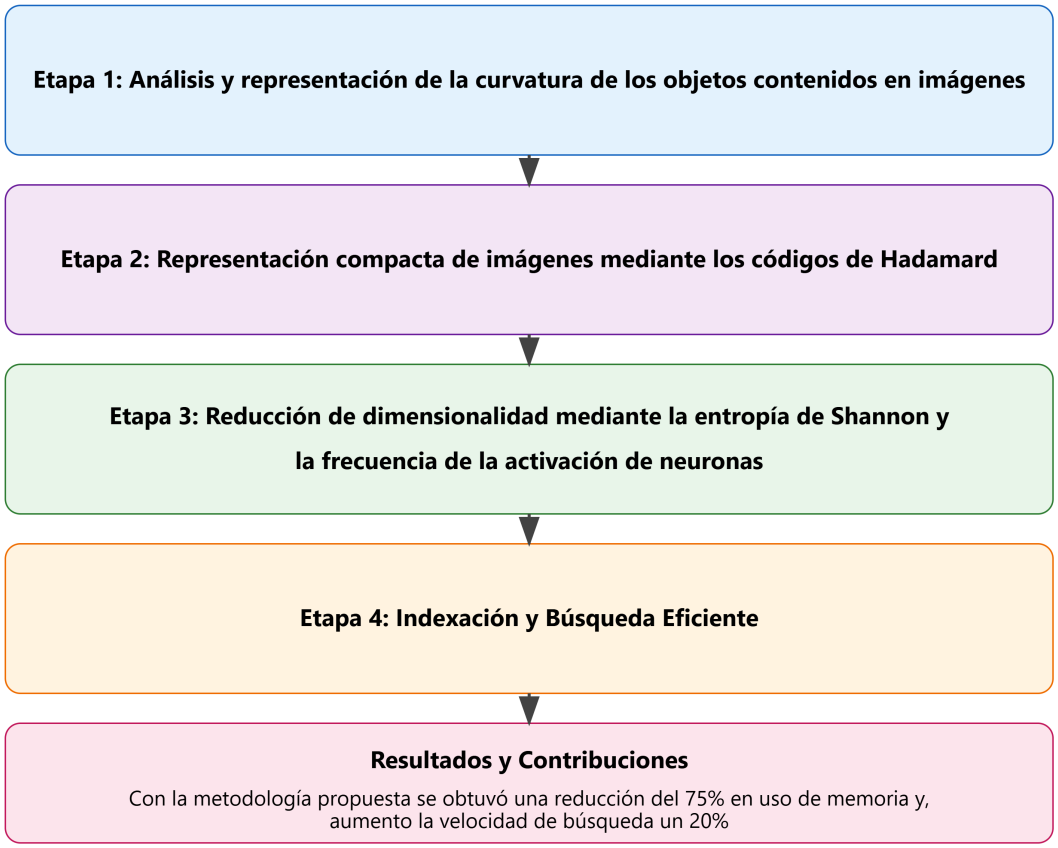


Figura 1.3: Metodología y contribuciones de la tesis

1.6. Metodología

La solución propuesta en esta tesis doctoral se desarrolla a través de una metodología estructurada en cuatro etapas fundamentales, cada una de ellas fue diseñada para abordar aspectos críticos del problema de recuperación de imágenes en grandes bases de datos, estas etapas se muestran en la Fig 1.3.

La primer etapa se enfoca al análisis y representación de la curvatura de los objetos, en esta etapa se desarrollaron técnicas de selección de puntos clave robustos, para crear patrones triangulares con ellos y de esta manera representar los objetos contenidos en imágenes. Esta representación fue validada ante diferentes transformaciones afines como lo son: rotación, escalamiento y deformación, obteniendo resultados experimentales favorables.

La segunda etapa trata de la representación compacta de imágenes mediante los códigos Hadamard. En esta etapa se proponen las características profundas de Hadamard (DHF) para reemplazar a las características profundas tradicionales. La evaluación comparativa entre este tipo de características demuestra la eficacia del enfoque propuesto.

La tercera etapa se enfoca en la reducción de dimensionalidad mediante la entropía de Shannon y la frecuencia de la activación de neuronas (TF-iDF). En esta etapa se desarrollaron diversos criterios para seleccionar neuronas valiosas. La validación del recall en espacios de baja dimensión confirman la efectividad de esta aproximación.

La cuarta etapa está enfocada a la indexación y búsqueda eficiente mediante las matrices de Hadamard como motor de búsqueda. La evaluación en esta etapa muestra mejoras significativas en cuanto a velocidad y recuperación de imágenes a la consulta.

La validación experimental se realizó utilizando bases de datos estándar como ImageNet [Deng09a], CIFAR-100 [Krizhevsky12a] y COCO [Lin14], junto con arquitecturas CNN establecidas como ResNet [He16a], VGG16 [Simonyan15] y EfficientNet [Tan19b]. Las métricas de evaluación incluyen Recall@k (k=1,5,10), tiempo de procesamiento, uso de memoria y robustez a transformaciones. La comparación con métodos del estado arte como FAISS [Douze24] y HNSW [Malkov20a] demuestra la superioridad de este trabajo.

Las contribuciones principales de esta investigación incluyen una nueva técnica de representación basada en curvatura, un método innovador de compresión mediante códigos

Hadamard, un algoritmo eficiente de selección de neuronas y un indexador rápido y robusto para grandes bases de datos.

Esta metodología integral sienta las bases para el desarrollo de sistemas CBIR eficientes y escalables, habilitando nuevas aplicaciones en dispositivos con recursos limitados, avances en el procesamiento de bases de datos grandes o big data y futuras investigaciones en visión computacional. Los resultados obtenidos demuestran el potencial de este enfoque para revolucionar el campo de la recuperación de imágenes.

1.7. Estructura de la tesis

En el Capítulo 2 son descritas diversas técnicas para recuperar objetos utilizando la curvatura de su forma. En el Capítulo 3 se propone el uso de códigos de Hadamard como alternativa a las características profundas para recuperar imágenes similares. En el Capítulo 4 se presentan técnicas para reducir el uso de memoria a través de la selección de neuronas valiosas de modelos CNNs. En el Capítulo 5 se introduce un nuevo indexador basado en las matrices de Hadamard para mejorar la velocidad en la recuperación de imágenes. Finalmente, se presenta las conclusiones de la investigación y se discute el trabajo futuro en el Capítulo 6.

Esta estructura permite una exploración sistemática de los desafíos y soluciones propuestas en el campo de la recuperación de imágenes basada en contenido, culminando en una metodología integral para ser aplicado en diferentes áreas.

1.8. Comentarios Finales

Este primer capítulo establece un marco sólido y crítico para abordar el desafío de la Recuperación de Imágenes Basada en Contenido (CBIR) en la era del big data. Por lo cual, fueron presentados antecedentes y el estado actual de la investigación en este campo, destacando los avances, y la necesidad de técnicas eficientes y escalables para la recuperación de imágenes en grandes bases de datos.

Como se vio anteriormente, la relevancia de CBIR trasciende el ámbito académico,

porque es una tecnología que puede ser implementada en una amplia gama de aplicaciones donde se requiera recuperar de imágenes similares. Por lo tanto, el verdadero reto radica en desarrollar técnicas que no solo sean muy precisas, sino que demuestren una escalabilidad robusta ante el crecimiento de datos, mientras mantienen una eficiencia computacional.

El equilibrio entre rapidez, eficiencia y escalabilidad representa el núcleo de la investigación presentada en esta tesis doctoral. La metodología propuesta en este trabajo no se limita a abordar estos desafíos de manera aislada, sino que está diseñada con técnicas puntuales para optimizar el uso de memoria, acelerar significativamente los tiempos de respuesta y mejorar la escalabilidad.

En los capítulos subsiguientes se desarrolla cada componente de esta metodología, y se demuestra cómo la sinergia entre las técnicas propuestas superan las limitaciones actuales, mientras se establecen nuevos estándares en el campo. Esta contribución representa un avance significativo en el estado del arte, porque proporciona una base sólida para el desarrollo de sistemas CBIR más robustos y eficientes, capaces de adaptarse al crecimiento continuo del big data.

En el siguiente capítulo se presenta un conjunto de técnicas fundamentales para la recuperación de imágenes basadas en la curvatura de la forma de los objetos. Estas técnicas no solo constituyen un punto de partida, sino que representan una solución inicial al desafío de la búsqueda de imágenes, usando solamente la curvatura de los objetos. El objetivo del siguiente capítulo está alineado perfectamente con el objetivo principal de esta tesis, porque, a través de estas técnicas, se pueden crear representaciones compactas y altamente informativas de imágenes.

Capítulo 2

Recuperación de Imágenes Basada en la Curvatura de su Forma

"La simplicidad es la máxima sofisticación."

Leonardo da Vinci (1452-1519), inventor y científico

La Recuperación de Imágenes Basada en Contenido (Content-Based Image Retrieval, CBIR) es un campo fundamental de visión computacional, donde se utilizan descriptores y las propiedades intrínsecas de los objetos, para identificar imágenes similares de grandes bases de datos. En el estado del arte, tres propiedades principales han dominado este campo: el color, que permite discriminar objetos por sus valores cromáticos; la textura, que captura patrones repetitivos y características superficiales; y la forma, que describe la estructura geométrica y los contornos de los objetos.

Entre estas propiedades, la forma se ha destacado como una de las características más importantes y distintivas por varias razones; En primer lugar, porque ofrece una representación robusta que permanece estable incluso cuando las condiciones de iluminación varían o el color del objeto cambia. Esta invariancia es valiosa en aplicaciones del mundo real, donde las condiciones de captura de imágenes pueden ser inconsistentes. Además, la forma tiende a ser más discriminativa que otras propiedades, permitiendo distinguir efi-

cazmente entre diferentes categorías de objetos, incluso cuando comparten características similares de color o textura.

La literatura sobre clasificación y recuperación de imágenes basada en forma es extensa y diversa, abarca desde métodos tradicionales basados en descriptores geométricos, hasta técnicas más recientes que emplean aprendizaje profundo. Los avances en este campo han llevado al desarrollo de descriptores sofisticados, capaces de capturar tanto características locales como globales de la forma, permitiendo de esta manera crear una representación más completa y precisa de los objetos.

En este capítulo se presenta una técnica innovadora para la recuperación de imágenes basada en la curvatura de los objetos. En esta propuesta son aprovechadas las ventajas inherentes de la forma como descriptor, mientras se abordan desafíos computacionales que han limitado la aplicación de técnicas similares en el big data. La técnica propuesta representa un avance significativo en esta área de interés, no solo por lograr un recall cercano a la perfección (0.99) tanto para objetos sin modificaciones, sino también por lograr un recall similar con objetos que han experimentado modificaciones como transformaciones afines, incluyendo escalamiento, rotación y deformaciones. Los resultados son particularmente notables en el caso de las deformaciones, donde la técnica propuesta demuestra su robustez y efectividad en comparación con el estado del arte.

2.1. Introducción

La Recuperación de Imágenes Basada en Contenido (Content-Based Image Retrieval, CBIR) es un campo fundamental de visión computacional donde se utilizan descriptores y las propiedades intrínsecas de los objetos, para reconocer imágenes similares a una consulta. Entre estas propiedades, la forma se destaca por su estabilidad y capacidad discriminativa, convirtiéndose en un pilar central de este campo de estudio [Yildirim21, Mokhtarian97, Zhang04].

Las representaciones de formas se dividen principalmente en dos categorías: basadas en regiones y basadas en contornos [Zhang03]. Cada una de estas se subdivide en métodos estructurales y globales, como se ilustra en la Figura 2.1.

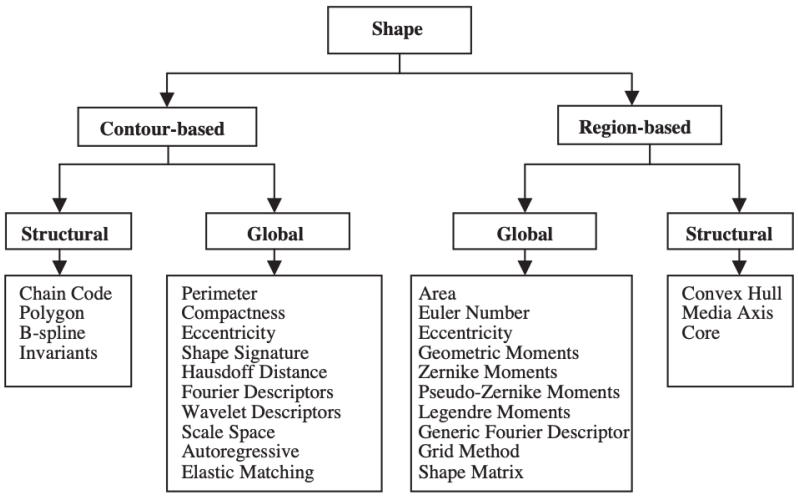


Figura 2.1: Taxonomía de la de representación de formas [Zhang04]

Los métodos estructurales, aunque son útiles para coincidencias parciales, presentan desafíos significativos, porque son sensibles al número de primitivas, y son computacionalmente más costosos [Zhang03]. Por otro lado, los métodos globales ofrecen mayor robustez frente a variaciones en la forma, haciéndolos más confiables para coincidencias completas [Zhang03, Yildirim21].

Una técnica popular en este campo de investigación, clasificada dentro de los métodos globales es la Normalización de Distancias Iguales (EDN). Con EDN se pueden diferenciar áreas representativas de los objetos, a través de muestras equidistantes [Paramarthalingam21a, Keogh09]. Este enfoque de muestreo ha sido muy utilizado en el estado del arte por su robustez, y su manera eficaz de obtener representaciones compactas y altamente informativas de los objetos. También ha sido la inspiración para diseñar e implementar la técnica propuesta en este capítulo de tesis doctoral.

Latif et al. [Latif19] han realizado un estudio comparativo de las técnicas de recuperación de imágenes basadas en forma, destacando la eficacia de cada método según el contexto de aplicación. Sus hallazgos subrayan la importancia de considerar múltiples factores al diseñar sistemas de recuperación, incluyendo la complejidad computacional, la robustez a transformaciones y la escalabilidad. Baroffio et al. [Baroffio16] al igual que Latif

et al. realizaron un análisis exhaustivo de la robustez de diferentes descriptores de forma, en su trabajo demostraron que la elección del descriptor no solo depende de la aplicación específica, sino también de las características de los datos, este análisis proporciona una base sólida para la selección de descriptores en aplicaciones prácticas. También, mostraron que la combinación de descriptores tanto globales y locales pueden mejorar significativamente la eficiencia en tareas de recuperación de imágenes, especialmente cuando se cuenta con oclusiones parciales [Mingqiang08].

Los avances recientes en el campo de la recuperación de imágenes basada en forma han experimentado una transformación significativa con la introducción de técnicas de aprendizaje profundo [Jiang24]. La integración de redes neuronales con descriptores está redefiniendo el estado del arte en este campo, este hecho no solo ha mejorado la precisión en la recuperación de imágenes, sino que también ha introducido nuevos desafíos en términos de eficiencia computacional y escalabilidad.

Para abordar la complejidad computacional inherente a estos desafíos, en las últimas décadas el uso de descriptores compactos e indexables se ha vuelto una práctica común [Paramarthalingam21a]. Así como el uso de estructuras de datos dinámicas como tablas Hash, árboles-kd, y árboles-M han ganado popularidad[Keogh09, Bartolini05].

Finalmente, la evaluación de los algoritmos CBIR es esencial para medir su eficacia. Zhang et al. [Zhang03, Zhang04] proponen seis factores clave: precisión de recuperación, compactación de características, generalidad de aplicación, baja complejidad computacional, robustez e incremento gradual de complejidad.

A pesar de los avances en esta área de interés, al momento de escribir este trabajo de investigación, la forma sigue siendo compleja de tratar, y al mismo tiempo es muy utilizada en visión computacional por su capacidad discriminante. Las técnicas propuestas, que utilizan solo la forma como descriptor, logran resultados asombrosos, sin embargo, son incapaces de alcanzar una tasa de recuperación perfecta utilizando únicamente este descriptor, por lo cual, sigue siendo un desafío [Alajlan07]. Este hecho ha llevado a muchos investigadores a enfocarse más en el contenido de la curvatura de la forma que en la forma en sí [Zhang04].

En este capítulo se presenta una técnica innovadora para la recuperación de imáge-

nes basada en la curvatura de la forma. Esta propuesta fusiona las ventajas de los métodos estructurales y globales, introduciendo puntos clave seleccionados estratégicamente para generar patrones triangulares. Este enfoque no solo ofrece robustez frente a transformaciones como rotación, escalamiento y deformación, sino que también facilita su indexación, para el manejo de grandes bases de datos de imágenes.

En la Figura 2.2 se ilustran las partes fundamentales de la técnica propuesta para extraer puntos clave sobre la forma de los objetos y su indexación. Esta figura se divide en cuatro secciones que muestran el proceso completo: En la sección a) se muestra una manzana a la cual se aplicará dicha técnica. Este objeto(manzana) es procesada inicialmente con el detector de bordes Canny para obtener su contorno, eliminando posibles agujeros en la forma. En la sección b) se visualizan los puntos clave seleccionados sobre el contorno de la manzana. En la sección c) se ilustran los triángulos generados a partir de los puntos clave seleccionados. Estos triángulos se construyen sistemáticamente, tomando uno de ellos como referencia a los puntos adyacentes de la izquierda y derecha hasta agotar todos los puntos clave disponibles, o en su defecto que no puedan ser generados más triángulos. En la sección d) se representan los números complejos asociados a cada triángulo, calculados mediante la función de transformación.

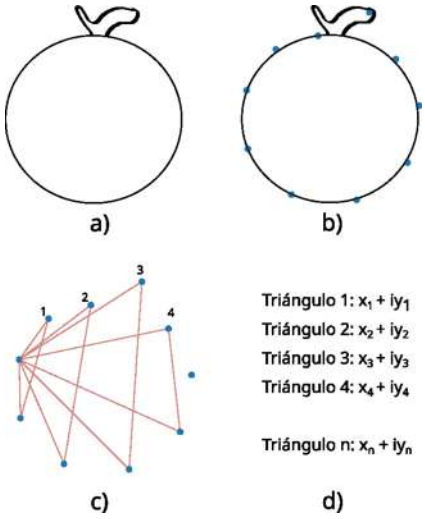


Figura 2.2: Diagrama de la técnica propuesta

Las contribuciones principales de este capítulo son:

- Una técnica novedosa para la representación y comparación de formas.
- Un método de indexación rápido y eficiente usando triángulos.
- Una técnica robusta a escalamiento, translación, rotación y cortes.

En este capítulo no solo se busca avanzar en el campo de la recuperación de imágenes basada en formas, sino también, proporcionar una base sólida para futuras aplicaciones en visión computacional y el reconocimiento de objetos usando las propiedades intrínsecas de las imágenes en el big data o bases de datos masivas.

Este trabajo fue presentado en la Conferencia Internacional Mexicana de Inteligencia Artificial (MICAI 2022) en el artículo "Recuperación de formas mediante coincidencia de polígonos".

2.2. Trabajo relacionado

Los primeros trabajos en el estado del arte acerca de la recuperación de imágenes usando la curvatura de la forma de los objetos, tratan acerca de las diferentes maneras posibles para crear secuencias de valores que representen curvaturas, ángulos, valores de descriptores o coeficientes poligonales de objetos. Uno de esos trabajos corresponde a la Curvatura del Espacio de Escala (Curvature Scale Space, CSS por sus siglas del inglés) propuesto por Mokhtarian et. al [Mokhtarian97]. En CSS se considera el cruce por cero de la curvatura de la forma para representar el contorno de los objetos, mediante cinco pares de valores. La desventaja de esa técnica es la sensibilidad al ruido, razón la cual los autores obtienen bajos resultados, su ventaja, es que es indexable [Mokhtarian97, Abbasi99]. A diferencia de CSS, la técnica propuesta utiliza una representación más robusta basada en triángulos, lo que la hace menos sensible al ruido, manteniendo su capacidad de ser indexable.

Otro descriptor utilizado por muchos investigadores es el descriptor de momentos de Zernike (Zernike Moment Descriptor, ZMD por sus siglas del inglés). Este descriptor tiene muchas propiedades deseables como invariancia de rotación, robustez al ruido y es

eficiente. Kim et. al. demostraron que ZMD se puede utilizar como un descriptor de forma global para buscar imágenes en grandes bases de datos [Kim00]. Los experimentos de estos investigadores fueron hechos con una base de datos de 6,000 imágenes aproximadamente, logrando resultados satisfactorios. Otros investigadores que han evaluado los descriptores ZMD y CSS, y que comparten la opinión con Kim con respecto a ZMD, son Zhang y Lu [Zhang03].

Kumar y Mali et. al. [Kumar21] destacan la importancia de crear secuencias de puntos clave desde el contorno, para clasificar objetos mediante el algoritmo Dynamic Time Warping (DTW, por sus siglas del inglés) y el algoritmo Dynamic Space Warping (DSW, por sus siglas del inglés). DTW y DSW han demostrado ser muy efectivos para medir la similitud de dos secuencias de datos en la literatura [Alajlan11, Yildirim21]. Sin embargo, estos algoritmos poseen dos dificultades asociadas; la primera está relacionada con secuencias de valores muy grandes, por el tiempo de cómputo requerido. La segunda dificultad surge cuando se pierde la referencia del punto inicial de la secuencia. DTW y DSW no están diseñados para elegir el punto inicial de las secuencias a comparar, este hecho empeora cuando se trabaja con objetos que han experimentado rotación en algún sentido. Para aminorar este efecto, Alajlan et al. proponen un algoritmo de comparación de cadenas similar al DTW, denominado como algoritmo HopDSW, cuyo fin es encontrar el punto de partida inicial de manera eficiente [Alajlan11].

Tomando en consideración este problema, en el desarrollo de la técnica propuesta se usa el centroide de los objetos, y el punto más cercano desde el contorno al centroide como referencias principales, con esta solución, no importa cuanto sean rotados los objetos, el inicio de la secuencia no se va a perder.

Por otro lado, Bartolini et. al. propusieron un enfoque de recuperación de formas llamado WARP, el cual está basado en la transformada de Fourier [Bartolini05]. Bartolini et. al. afirman que la información de fase proporciona una descripción más precisa de los límites del objeto a solo usar los coeficientes de Fourier, al igual que Kumar y Mali utilizan el algoritmo DTW para comparar curvaturas, además implementan índices de proximidad para la recuperación de imágenes.

La idea de obtener patrones a partir de la curvatura de la forma de los objetos

llevó a muchos investigadores a generar triángulos sobre ella [Alajlan07]. La elección de triángulos se ha sobrepuesto a otros polígonos, por su versatilidad para representar objetos. En aras de resaltar este hecho, Alajlan et. al. [Alajlan07] propusieron TAR (Triangle-Area Representation, TAR por sus siglas del inglés). TAR es un algoritmo de recuperación de imágenes a través de triángulos creados con puntos clave, sobre la curvatura de los objetos. En sus resultados muestran que su algoritmo es muy efectivo y excelente en condiciones donde los objetos a reconocer poseen contornos cerrados, y sin agujeros, en otro caso tienden a presentar problemas de reconocimiento. La principal ventaja de TAR, es su eficacia para capturar características tanto locales como globales de las formas; además de ser invariante a la traslación, la rotación y el escalado, es resistente al ruido y a algunas oclusiones parciales. En la etapa de comparación, Alajlan et. al. utilizan el algoritmo DSW para buscar la correspondencia entre los puntos de dos formas y calculan la distancia en función de la mejor alineación entre dos representaciones de formas.

La técnica propuesta de este capítulo de tesis está alineada con los objetivos de Alajlan et. al. [Alajlan07] respeto a crear representaciones invariantes a transformadas afines como translación, rotación y escalado. La diferencia consiste en que TAR requiere contornos cerrados y sin agujeros, a diferencia de la técnica propuesta, la cual es más flexible en este aspecto, además, por utilizar el centroide como punto de referencia fijo, en la propuesta se ofrece una mayor eficiencia computacional para manejar objetos ocluidos.

Keogh et. al. consideran a la rotación como un reto muy complejo, por tal motivo, estos autores han enfocado su trabajo a la implementación de diferentes técnicas para rotar objetos minimizando la degradación del recall[Keogh09].

Yildirim et. al. propusieron un enfoque estadístico en [Yildirim21], donde calculan la desviación estándar desde contorno y los ángulos de la forma al centroide. Estos autores cuantizan los ángulos a valores enteros, luego, para cada ángulo, extraen tres características que son; (1) el número de repeticiones de contorno; (2) la distancia promedio de los puntos en ese ángulo al centroide; y (3) la desviación estándar de esas distancias. Mientras que Yildirim et al. se enfocan en la desviación estándar de los ángulos desde el centroide, en la técnica propuesta se extiende este concepto, al utilizar ángulos para generar una representación triangular, con la cual se capturan mejor las características estructurales de la forma.

Kumar y Mali utilizan el centroide como punto fijo para trazar líneas perpendicular desde cada punto del contorno del objeto, hasta pasar por el punto fijo. En la etapa de comparación, estos autores utilizan análisis de componentes principales (PCA) de las distancias perpendiculares obtenidas; su método es robusto a traslaciones y rotaciones [Kumar21]. La técnica propuesta comparte similitudes con el enfoque de Kumar y Mali en el uso del centro de gravedad como punto de referencia, pero se diferencia con el uso de patrones triangulares y la transformación a números complejos para su indexación

Xu et. al. implementaron un método de recuperación llamado Corner-Guided DP, el cual se enfocan en la recuperación de imágenes mediante coincidencia parcial de objetos (Partial Shape Matching, PSM por sus siglas del inglés). Corner-Guided DP fue hecho con el objetivo de reconocer eficazmente los huesos humanos sobre imágenes de rayos X. El funcionamiento de este método se basa en nueve puntos de referencia para crear múltiples triángulos. Este enfoque es eficiente, rápido, y robusto a transformaciones afines como traslación, rotación y escala [Xu08]. Similar a Corner-Guided, en la técnica propuesta se utilizan puntos de referencia para generar triángulos. A diferencia con ese método, respecto a los nueve puntos fijos, en la técnica propuesta se permite una selección más flexible para adaptarse mejor a diferentes tipos de formas.

Arjun y Mirnalinee propusieron un algoritmo iterativo llamado Integración de Características Multiescala (ICM) el cual utiliza puntos de la curvatura de la forma; estos puntos se ordenan de acuerdo con su distancia normalizada al contorno. Para la extracción de características Arjun y Mirnalinee, utilizan algoritmos de patrón angular como lo son: AP, AP binario (BAP) y selección secuencial hacia atrás (SBS) [Arjun18].

Abro et. al. [Abro19] evaluaron varios descriptores con el objetivo de mostrar que, los resultados pueden mejoran usando varios descriptores al mismo tiempo. Los descriptores evaluados por estos autores corresponden a descriptores de Fourier, centroides jerárquicos, descriptores basados en momentos y descriptores de contexto de forma. La precisión que lograda en su trabajo corresponde a un 90 %.

Paramarthalingam y Thankanadar et al. propusieron un procedimiento para generar puntos normalizados a partir de siluetas de formas. Este procedimiento identifica el contorno de cualquier objeto en una imagen y utiliza el método de Normalización del Área

del Objeto (Object Area Normalization, OAN por sus siglas del inglés) para dividir el objeto por su centro en regiones con la misma área. Estos autores definieron seis descriptores para modelar la forma del objeto [Paramarthalingam21b]; Distancia Centroide Compacta (CCD), Consejo Central (ANG), Distancia de Puntos Normalizados (NPD), Relación de Distancia Centroide (CDR), Descriptor de Patrón Angular (APD), y Representación de Área Multitriángulo (MTAR). Similar a OAN, donde se buscan regiones de igual área. En la técnica propuesta son creados intervalos angulares para seleccionar puntos clave de las formas. Esto permite una representación robusta a deformaciones mientras mantiene la eficiencia computacional.

Zhang et. al. proponen Shape Classification Network (SCN por sus siglas del inglés). SCN es una técnica basada en el modelo convolucional LeNet5, para reconocer imágenes con números escritos a mano. Estos autores muestran una alternativa interesante a crear patrones sobre la curvatura de los objetos, en su lugar entrenan modelos convolucionales para reconocer los trazos de las formas de objetos [Zhang21]. Este enfoque abre nuevos panoramas a esta área de interes usando modelos neuronales.

Damen et. al. utilizan constelaciones de puntos robustos para detectar imágenes en una transmisión de vídeo [Damen12]. Esta técnica, aunque pareciera sencilla, tuvo un gran impacto en el reconocimiento de imágenes, no solo porque es muy práctica para caracterizar formas aisladas correctamente, también funciona con objetos parcialmente ocluidos. La idea de utilizar constelaciones, también fue inspiración para la técnica propuesta en este trabajo de investigación, con unas ligeras diferencias, en lugar de usar constelaciones de puntos, se utilizan patrones geométricos para caracterizar formas. La ventaja de la técnica propuesta sobre el uso de constelaciones, es la representación compacta y computacionalmente eficiente que se obtiene de los triángulos, al cual facilita su indexación.

En la literatura existe una amplia variedad de trabajos especializados en esta área de interés, en este capítulo, solo han sido considerados aquellos cuya precisión/recall oscila entre el 90 % y el 100 % para rotación o escalamiento, desafortunadamente, no es común para ambos casos. Por otro lado, existe una transformación cuya complejidad no es abordada por muchos investigadores, este es el caso de la transformación que deforma las imágenes, conocida como corte (Shear). En la literatura es difícil encontrar trabajos que la tomen en

consideración, y los pocos autores que la incluyen sus resultados, no superan el 80-85 % en precisión.

La técnica propuesta está inspirada en algunas fortalezas de trabajos previos, como la manera de seleccionar puntos clave robustos desde la curvatura, los descriptores de fourier y algunas estrategias de indexación. Por su parte, tiene ventajas competitivas por:

- Permitir una selección flexible de puntos clave basada en intervalos angulares desde el centroide, los cuales proporcionan invariancia a rotación y escalamiento.
- Crear una representación mediante triángulos, con la cual se capturan eficientemente tanto características locales como globales de la forma, y es robusta a deformaciones y oclusiones parciales.
- Implementar una transformada de triángulos a números complejos indexable.
- No requerir entrenamiento previo ni grandes recursos computacionales.
- Gestionar de manera natural formas irregulares y no cerradas.

Estas características superan las limitaciones de las técnicas previas, por lo tanto, a través de la técnica propuesta, son sentadas las bases para sistemas de recuperación versátiles, robustos y eficientes.

Antes de finalizar esta sección, es conveniente mencionar que información básica de este tema se encuentra disponible en Jain et. al. [Jain96] y Zang et. al. [Zhang04], por su parte Yildirim et. al. [Yildirim21] proporcionan un resumen con los avances más recientes respecto a este tema.

2.3. Técnica propuesta

La técnica propuesta se basa en la representación de la curvatura de la forma, mediante un conjunto de triángulos creados a partir de puntos clave seleccionados estratégicamente. Este enfoque combina las ventajas de los métodos estructurales y globales, ofreciendo una representación compacta de la curvatura del objeto en cuestión.

Los métodos estructurales permiten capturar detalles locales significativos a través de la selección de puntos clave, mientras que el enfoque global, asegura la robustez frente a transformaciones como rotación, escalamiento y deformaciones moderadas.

La selección de puntos clave que toman en consideración criterios geométricos garantiza que los polígonos resultantes capturen las características esenciales de la forma, mientras mantienen una representación computacionalmente eficiente. Por otro lado, la transformación de estos triángulos a números complejos, preserva las relaciones geométricas importantes, mientras facilita su indexación y recuperación en grandes bases de datos.

2.3.1. Fundamentos teóricos

La técnica propuesta se fundamenta en los siguientes tres principios teóricos:

El primer principio se basa en la teoría de la forma y la curvatura. La curvatura actúa como un descriptor invariante a transformaciones afines, permitiendo caracterizar objetos independientemente de su orientación o escala, como lo demuestran Yildirim et al. [Yildirim21] y Zhang et al. [Zhang04] en sus estudios sobre descriptores de forma.

El segundo principio se centra en el análisis geométrico, particularmente en las propiedades de los triángulos como descriptores de forma. Los triángulos ofrecen una representación robusta a transformaciones afines, como lo establecen Alajlan et al. [Alajlan07] en su trabajo sobre Triangle-Area Representation (TAR). Este principio se refuerza con los hallazgos de Paramarthalingam y Thankanadar [Paramarthalingam21a], quienes demuestran la efectividad de los triángulos para generar descriptores compactos.

El tercer principio se basa en las garantías matemáticas descritas por Chávez et al. [Chávez13] para la comparación de polígonos mediante números complejos. La estabilidad del método frente a pequeñas perturbaciones y su capacidad para preservar características esenciales han sido validadas también por Zhang et al. [Zhang03] en su evaluación exhaustiva de métodos de representación de formas. Estos fundamentos convergen para crear una técnica que es invariante a escalamiento, rotación, y traslación, mientras mantiene robustez frente a deformaciones moderadas. Como demuestran Keogh et al. [Keogh09] y Bartolini et al. [Bartolini05], estas propiedades son esenciales para sistemas eficientes de recuperación de imágenes basados en forma.

2.3.2. Comparación de polígonos

En Chávez et al. proponen un método para comparar polígonos, cuya esencia está relacionada con el descriptor de Fourier, el cual es invariante a transformaciones afines como rotación, traslación, escalado y deformaciones [Chávez13]. En ese trabajo, se sabe que un polígono es un conjunto ordenado de puntos consecutivos, o bien, un conjunto ordenado de números complejos ($\mathbb{C}$), cada uno a razón de $z = x + jy$, donde $j = \sqrt{-1}$. La comparación de polígonos está diseñada a manera que son comparadas secuencias de puntos o vértices, también vistos como secuencias de números complejos. Si dos polígonos están relacionados por afinidad, entonces la distancia entre sus valores asociados son una constante [Hernández17, Chávez13, Chávez16].

El enfoque de Chávez [Chávez13] para lograr la similitud de polígonos implica la construcción de funciones escalares complejas del tipo: $\varphi_\ell : \mathbb{C}^n \to \mathbb{C}, \ell = 1, \cdots, \lfloor (n-1)/2 \rfloor$, donde n es el número de vértices (3 para triángulos). Es importante mencionar que todos los polígonos similares que han sufrido la misma transformación, van a ser representados por el mismo número complejo (φ_ℓ), donde φ_ℓ se obtiene a partir de la siguiente función $\varphi_\ell : \mathbb{C}^n \to \mathbb{C} \bigcup \{\infty\}$ de acuerdo a la siguiente Ecuación:

$$\varphi_\ell(z_1, z_2, z_3, \cdots, z_n) = \frac{\sum_{k=1}^{n} \lambda^{\ell k} z_k}{\sum_{k=1}^{n} \lambda^{-\ell k} z_k} \tag{2.1}$$

En la Ecuación 2.1 $\lambda = e^{\frac{2\pi j}{n}}$ corresponde a la n-ésima raíz unitaria y z_k son los vértices del triángulo representados como números complejos. Es conveniente mencionar que asignar a cada polígono un número complejo con la Ecuación 2.1, en la mayoría de los casos es un proceso rápido, pero depende mucho del equipo de cómputo. A diferencia de encontrar los polígonos similares en una colección, esta tarea es muy veloz.

Para finalizar, es importante mencionar que en este capítulo se ha explorado específicamente la generación y uso de triángulos para la representación de formas, aprovechando sus propiedades matemáticas y su simplicidad estructural. Sin embargo, los principios y técnicas desarrollados podrían extenderse a polígonos de mayor número de lados. El uso de cuadrados, pentágonos u otros polígonos regulares podría ofrecer representaciones alternativas con diferentes características de robustez y precisión. Esta extensión a otros

polígonos representa una dirección prometedora para investigaciones futuras, especialmente en casos donde se requiera capturar características más complejas de las formas.

2.4. Implementación de la propuesta

La implementación de la técnica propuesta se divide en tres etapas principales como se muestra en la Figura 2.3.

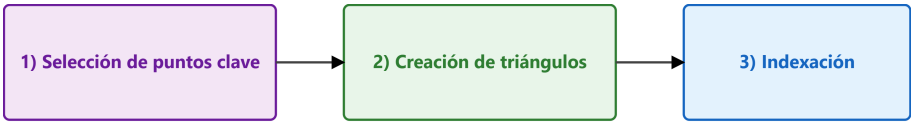


Figura 2.3: Etapas de la técnica propuesta

1. **Selección de puntos clave**: Se identifican puntos clave sobre el contorno del objeto usando el centroide como referencia y el primer punto más cercano del contorno a él.

2. **Creación de triángulos**: Los triángulos se generan sistemáticamente a partir de los puntos clave seleccionados.

3. **Indexación**: Cada triángulo se transforma en un número complejo y se almacena en una tabla hash de tamaño 256.

A continuación se explica en profundidad en que consisten estas etapas:

2.4.1. Selección de puntos clave

Para seleccionar puntos clave en la técnica propuesta, se sigue el siguiente proceso:

1. Determinar la forma del contorno de la imagen, para eso se utiliza el detector de bordes Canny, sin olvidar eliminar los agujeros de la figura con el método de erosión, similar a los que se muestran en el ciervo de la Figura 2.4.

2. Se determina el centroide de la forma del contorno y el punto más cercano a este centroide que se encuentra en el contorno. Estos dos puntos clave, etiquetados como

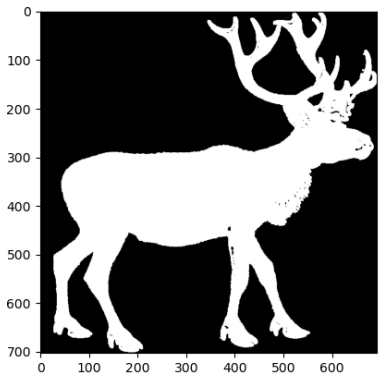


Figura 2.4: Figura de ciervo con agujeros

1 y 2, definen una línea de referencia. El punto 1 corresponde al centroide y el punto 2 es el punto más cercano del centroide.

3. Se traslada la imagen a modos que el centroide corresponda con el origen (0, 0).

4. A partir de la línea de referencia, son seleccionados puntos equidistantes en intervalos angulares específicos, estos intervalos son opcionales, por ejemplo, cada 120°, 90°, 72°, 60°, 51.42°, 45°, 40° y 36°, en sentido antihorario con respecto a las referencias principales. Con esto se obtienen desde 3 hasta 10 puntos de referencia respectivamente. Como este parámetro es opcional, en caso de requerir más muestras se debe de calcular el intervalo angular respectivo.

   Se recomienda etiquetar los puntos clave seleccionados, en el mismo orden en que son obtenidos, esto con el fin de tener puntos consecutivos.

En la Figura 2.5, se muestra el resultado tras aplicar los pasos previos a la figura del ciervo original (Figura 2.4). En esta Figura 2.5 han sido seleccionadas 6 muestras (cada 60°). En esta imagen se puede apreciar el centroide del ciervo (punto 1), así como cada muestra numerada en sentido antihorario (2-7).

2.4.2. Creación de triángulos

Para la creación de triángulos en la técnica propuesta, se sigue el siguiente proceso:

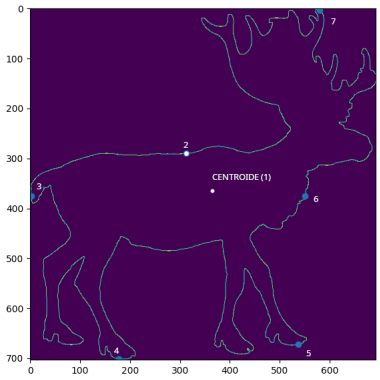


Figura 2.5: Imagen de ciervo con 6 muestras seleccionadas (cada 60 grados)

- Los triángulos se forman uniendo el punto 2 de referencia a los puntos adyacentes (izquierda y derecha)
- El proceso continúa hasta agotar los puntos clave disponibles o hasta que no haya suficientes puntos para formar un nuevo triángulo

2.4.3. Indexación

En la indexación y recuperación de triángulos, se sigue el siguiente proceso:

- Por cada triángulo construido, se aplica la Ecuación 2.1 para obtener el número complejo relacionado, posteriormente se calcula su magnitud, con eso se obtiene un solo número por triángulo.
- Son almacenados tanto el número obtenido del paso anterior del triángulo, así como su identificador único en una tabla Hash de tamaño 256. La forma tiene una entrada a la tabla Hash para cada triángulo construido a partir de ella.

Para buscar imágenes similares, se procesa la consulta a modo que son obtenidos los triángulos del objeto hasta los códigos Hash relacionados. Con estos códigos se buscan en la tabla Hash la mejor coincidencia. Los objetos que comparten más triángulos similares se consideran más parecidos a la consulta. Este proceso es eficiente porque aprovecha la tabla Hash para realizar búsquedas rápidas usando los números calculados para cada triángulo como claves de búsqueda.

El Algoritmo 1 muestra la implementación de la técnica propuesta. En este Algoritmo cada etapa está diseñada para abordar un aspecto específico de la técnica propuesta.

La Fase 1, correspondiente a la selección de puntos clave (líneas 2-16), comienza con el procesamiento de la imagen de entrada utilizando el detector de bordes Canny, seguido de la eliminación de agujeros mediante erosión para obtener un contorno limpio. Posteriormente, se calcula el centroide (p_1) y se identifica el punto más cercano del contorno a este centroide (p_2). El contorno se traslada para que el centroide coincida con el origen del sistema de coordenadas. El proceso continúa con una iteración donde se proyectan líneas desde el centroide en intervalos angulares específicos, se identifican las intersecciones con el contorno y se selecciona el punto más externo cuando existen múltiples intersecciones. Todos estos puntos clave seleccionados se almacenan para su uso posterior.

La Fase 2, se enfoca en la creación de triángulos (líneas 18-24). Esta etapa inicia con la inicialización de un conjunto vacío destinado a almacenar los triángulos. En la construcción de los triángulos se utiliza el punto de referencia p_2 como vértice común y conectándolo con pares consecutivos de puntos clave seleccionados en la fase anterior. Este proceso asegura una representación completa de la forma del objeto.

La Fase 3, aborda la indexación (líneas 25-30) y comienza con la creación de una tabla Hash de tamaño 256. Para cada triángulo construido en la fase anterior, se calcula su número complejo utilizando la Ecuación 2.1. El valor resultante, junto con el identificador del triángulo, se almacena en la tabla hash, creando así una estructura eficiente para la búsqueda posterior.

La efectividad de este algoritmo se fundamenta en su capacidad para mantener la invariancia ante transformaciones afines mientras proporciona un mecanismo de búsqueda eficiente mediante tablas Hash. La selección cuidadosa de puntos clave y la construcción sistemática de triángulos garantizan una representación robusta de la forma, permitiendo identificar similitudes incluso cuando las imágenes han sufrido transformaciones afines.

Algoritmo 1 Recuperación de imágenes basada en forma mediante triángulos

Entrada: Imagen de entrada I, ángulo de muestreo θ

Salida: Triángulos indexados en tabla hash A

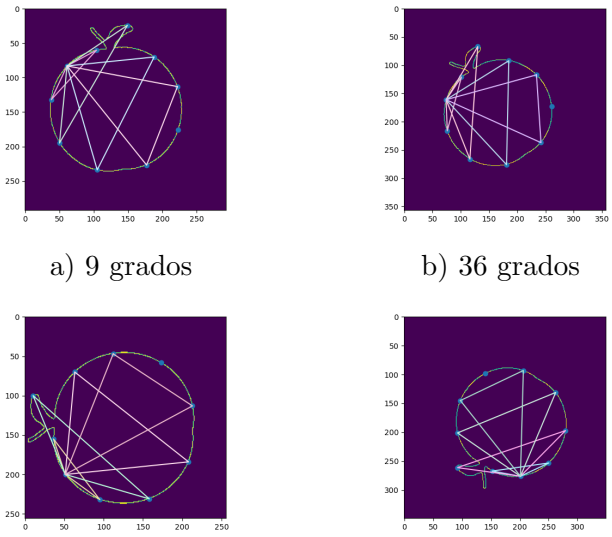
Fase 1: Selección de puntos clave
$I \leftarrow$ detectarContorno(I) usando Canny
$I \leftarrow$ eliminarAgujeros(I) usando erosión
$p_1 \leftarrow$ calcularCentroide(I)
$p_2 \leftarrow$ puntoMásCercano(I, p_1)
trasladar(I, $-p_1$) {Centrar en origen}
$P \leftarrow \{p_1, p_2\}$ {Conjunto de puntos clave}
$\alpha \leftarrow 0$ {Ángulo inicial}
mientras $\alpha < 360$ **hacer**
    $l \leftarrow$ trazarLinea(p_1, α)
    $intersecciones \leftarrow$ encontrarIntersecciones(l, I)
    si $|intersecciones| > 0$ **entonces**
        $p \leftarrow$ puntoMasExterno($intersecciones$)
        $P \leftarrow P \cup \{p\}$
    fin si
    $\alpha \leftarrow \alpha + \theta$
fin mientras
Fase 2: Creación de triángulos
$T \leftarrow \emptyset$ {Conjunto de triángulos}
$n_c \leftarrow |P|$ {Número de puntos clave}
para $i \leftarrow 3$ to n_c **hacer**
    $t \leftarrow$ Triángulo(p_2, $P[i]$, $P[(i \bmod n_c) + 1]$)
    $T \leftarrow T \cup \{t\}$
fin para
Fase 3: Indexación
$A \leftarrow$ crearTablaHash(256)
para cada triángulo $t \in T$ **hacer**
    $poligono \leftarrow$ calcularPoligonos(t) {Usando Ec. 2.1}
    agregarAHash(A, $poligono$, t)
fin para
devolver A

2.4.4. Criterios de implementación

La implementación de la técnica propuesta debe considerar varios aspectos críticos para asegurar su efectividad:

1. **Selección de puntos clave**: La robustez de los puntos clave es fundamental para la técnica propuesta. El primer punto seleccionado es especialmente crítico, ya que los triángulos subsecuentes se construyen a partir de esta referencia. Por esta razón, se utiliza el punto más cercano al centroide como punto inicial, ya que este tiende a ser más estable frente a transformaciones. Esta estrategia asegura que incluso cuando la forma experimenta rotaciones, escalamientos o deformaciones moderadas, el punto inicial sigue siendo identificable, proporcionando una base consistente para la construcción de triángulos subsecuentes.

   La Fig. 2.6 muestra la importancia de una buena selección de puntos clave usando la técnica propuesta en este capítulo de tesis, aunque la imagen de referencia sea rotada o escalada, aun así van a ser seleccionados los mismos puntos clave o en su caso un punto muy cercanos al original una variación pequeña que oscila entre los 1-3 píxeles.



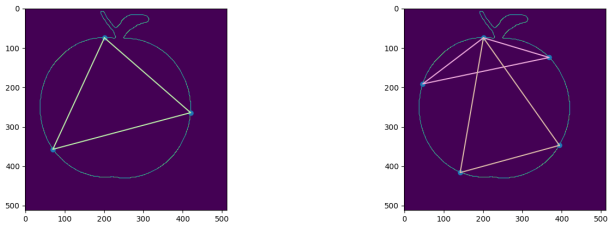
a) 9 grados b) 36 grados

c) 90 grados d) 150 grados

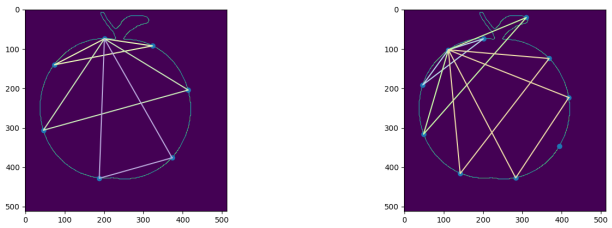
Figura 2.6: Manzana rotada a 9, 36, 90 y 150 grados

2. **Densidad de puntos clave**: La cantidad de puntos clave seleccionados impacta directamente en el número y tamaño de los triángulos generados. Un número reducido de puntos produce menos triángulos, pero de mayor tamaño, lo que puede resultar en una representación más robusta pero menos detallada. Por el contrario, un mayor número de puntos genera más triángulos de menor tamaño, capturando más detalles pero potencialmente introduciendo sensibilidad al ruido. La selección óptima depende de las características específicas del conjunto de datos y los requisitos de la aplicación.

La Figura 2.7 muestra una manzana representada con diferentes conjuntos de triángulos, esto con el fin brindar un ejemplo acerca del problema relacionado con la generación de puntos clave. En este contexto, no existe un número mínimo o máximo de puntos que se deben crear, este parámetro depende de las consideraciones del investigador. Sin embargo, es fácil de observar en esta Figura 2.7, que al variar el número de puntos clave, la cantidad de triángulos va a aumentar o disminuir, lo que permite expandir o acotar las características de referencia, como se muestra en esta figura con 3, 5, 7 y 10 puntos clave.



a) 3 puntos b) 5 puntos



c) 7 puntos d) 10 puntos

Figura 2.7: Selección de puntos clave en diferentes configuraciones

3. **Cruces múltiples del contorno**: Cuando una línea de referencia angular intercepta el contorno en múltiples puntos (por ejemplo, en formas complejas o con concavidades), se debe establecer un criterio de selección. En estos casos, se opta por seleccionar el punto más externo, favoreciendo la creación de triángulos más grandes. Esta decisión se basa en que los triángulos más grandes tienden a ser más estables y proporcionan una representación más robusta de la forma global del objeto.

   En la Figura 2.8 los puntos rojos resaltan las intercepciones, entre las referencias angulares, con la forma del objeto. En estos casos se opta por los puntos que se encuentren más exterior, como se puede ver con los puntos azules.

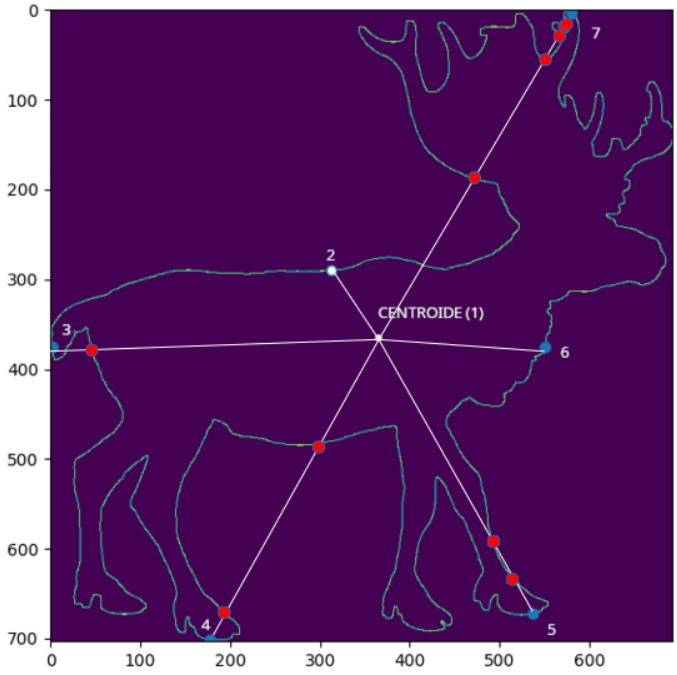


Figura 2.8: Imagen de ciervo con trazos desde el centroide al exterior

Estos criterios no solo afectan el recall del método, sino que también influyen en su aplicabilidad a diferentes escenarios y tipos de formas. Es importante ajustar estos parámetros según las características específicas del problema y la naturaleza de las formas a analizar.

2.5. Resultados experimentales

Los experimentos se diseñaron siguiendo un protocolo riguroso para garantizar la reproducibilidad de los resultados. Para evaluar la técnica propuesta se utilizó el conjunto de datos MPEG-7 Core Experiment CE-Shape-1, una base que consta de 1,400 imágenes distribuidas uniformemente en 70 clases de objetos diferentes [Latecki00]. Este conjunto se seleccionó por su amplia adopción en la comunidad científica y su diversidad en formas y características.

En el estado del arte están establecidos dos casos de estudio específicos para realizar las pruebas:

- **Caso I o Experimento normal**: Las imágenes se utilizan directamente sin modificaciones, preservando todas sus características originales.
- **Caso II o Experimento extendido**: Las imágenes son sometidas a múltiples variaciones como escalamiento, rotación y cortes, siguiendo los parámetros establecidos en Kimia [Kim00].

En ambos casos de estudio se procesan todos los triángulos obtenidos por cada imagen de consulta para buscar coincidencias utilizando la tabla Hash como estructura de indexación principal.

2.5.1. Experimento normal

En este experimento se implementó un incremento gradual en la cantidad de puntos de referencia, comenzando desde 3 hasta alcanzar 10 puntos. Es importante destacar que 3 puntos constituyen la unidad mínima para formar un triángulo, mientras que 10 puntos permiten la construcción de al menos 3 triángulos. Este rango se seleccionó considerando:

- La necesidad de mantener un equilibrio entre complejidad computacional y precisión
- La capacidad de capturar características discriminativas suficientes
- La minimización del ruido en el proceso de extracción de características

Los resultados experimentales demuestran que la técnica propuesta alcanza una precisión notable de 0.9957 de manera consistente (1,394 de 1,400 imágenes correctamente reconocidas). La Tabla 2.1 presenta los resultados detallados, donde se puede observar que el rendimiento se mantiene estable independientemente del número de puntos de referencia utilizados.

Nº Puntos	Aciertos/Total	Recall
3	1394/1400	0.9957
4	1394/1400	0.9957
5	1394/1400	0.9957
6	1394/1400	0.9957
7	1394/1400	0.9957
8	1394/1400	0.9957
9	1394/1400	0.9957
10	1394/1400	0.9957

Tabla 2.1: Resultados detallados del **Experimento normal**, mostrando la consistencia en el rendimiento independientemente del número de puntos de referencia

La consistencia observada en este experimento merece un análisis detallado. Este comportamiento puede atribuirse a varios factores clave:

1. **Estabilidad del descriptor triangular**: El uso de triángulos como unidad básica proporciona una representación inherentemente estable. Esto se debe a que:
   - Las relaciones geométricas entre los tres puntos que forman un triángulo son invariantes bajo transformaciones de similitud
   - Los triángulos capturan eficientemente las características locales de forma, independientemente del número total de puntos de referencia
2. **Saturación de información**: La consistencia en los resultados sugiere que tres puntos de referencia (un triángulo) ya proporcionan información suficientemente discriminativa para la tarea de reconocimiento. El incremento en el número de puntos no mejora significativamente el rendimiento porque:
   - La información adicional puede ser redundante con la ya capturada por el primer triángulo

- Las características esenciales que diferencian las clases están presentes en la configuración más simple

3. **Robustez del método de indexación**: La tabla Hash utilizada para el almacenamiento y búsqueda de características demuestra ser:
   - Altamente efectiva en la preservación de la información discriminativa
   - Resistente al ruido introducido por variaciones en el número de puntos
   - Capaz de mantener un rendimiento constante independientemente de la cantidad de datos indexados
4. **Naturaleza de los errores persistentes**: Las 6 imágenes no reconocidas correctamente (1400 - 1394) probablemente representan casos donde:
   - Existen ambigüedades inherentes en la forma que no pueden resolverse solo con información geométrica local
   - Las características distintivas de la clase no están bien representadas por la configuración triangular
   - Hay presencia de ruido o variaciones que afectan la estabilidad de los puntos clave seleccionados

Esta consistencia en el rendimiento tiene importantes implicaciones prácticas:

- **Eficiencia computacional**: Se puede obtener un rendimiento óptimo utilizando el mínimo número de puntos (3), reduciendo significativamente el costo computacional.
- **Escalabilidad**: El método puede aplicarse eficientemente a grandes conjuntos de datos sin necesidad de aumentar el número de puntos de referencia.
- **Robustez**: La estabilidad en el rendimiento sugiere que el método es robusto frente a variaciones en la selección de puntos clave.

Sin embargo, es importante señalar que esta consistencia también plantea interrogantes sobre posibles mejoras futuras. El hecho de que aumentar el número de puntos no mejore

el rendimiento sugiere que para superar el límite actual de precisión (0.9957) podría ser necesario:

- Incorporar información adicional más allá de la geometría triangular
- Desarrollar estrategias específicas para manejar los casos de fallo persistentes
- Explorar métodos complementarios que puedan resolver las ambigüedades remanentes

2.5.2. Experimento extendido

El experimento extendido se diseñó para evaluar la robustez del método frente a diversas transformaciones geométricas. Al igual que en el experimento normal, se incrementó gradualmente la cantidad de puntos de referencia de 3 a 10, pero en este caso se aplicaron transformaciones a las imágenes.
Los conjuntos de datos especializados se categorizaron como A1 y A2:

- **A1 - Evaluación de escalamiento**: Este conjunto se enfoca en evaluar la invarianza del método frente a cambios de escala. Las imágenes presentan diferentes escalas del mismo objeto, permitiendo verificar si el algoritmo mantiene su capacidad de reconocimiento independientemente del tamaño. Esta prueba es necesaria en aplicaciones donde la distancia al objeto puede variar.
- **A2 - Evaluación de rotación**: Este conjunto está diseñado específicamente para evaluar la invarianza rotacional del método. Las imágenes contienen el mismo objeto en diferentes orientaciones, simulando condiciones reales donde los objetos pueden aparecer en cualquier ángulo.

Las modificaciones aplicadas siguieron parámetros estandarizados, utilizados previamente en conjuntos de datos reconocidos como Kimia99 y ETH-80:

- **Escalamiento**: Se aplicaron factores de 0.1, 0.2, 0.25, 0.3 y 2.0, cubriendo tanto reducciones significativas como ampliaciones moderadas.
- **Rotación**: Se evaluaron ángulos de 9^{o}, 36^{o}, 45^{o}, 90^{o} y 150^{o}, abarcando tanto rotaciones sutiles como transformaciones más dramáticas.

- **Deformación**: Se aplicaron factores de -0.3, -0.2, -0.1, 0, 0.1 y 0.2, simulando distorsiones que pueden ocurrir en escenarios reales.

Los resultados del experimento extendido se presentan en la Tabla 2.2. Los hallazgos más significativos incluyen:

- Para el **escalamiento**, los mejores resultados (1399/1400) se obtuvieron utilizando entre 7 y 9 puntos clave
- En **rotación**, el rendimiento óptimo (1394/1400) se alcanzó con 3-4 puntos clave
- Para **deformación**, se mantuvo un rendimiento consistente (1394/1400) independientemente del número de puntos

Nº Points	Escalamiento	Rotación	Deformación	Promedio
3	1388/1400	**1394/1400**	1394/1400	**0.9942**
4	1382/1400	**1394/1400**	1394/1400	0.9928
5	1397/1400	1379/1400	1394/1400	0.9928
6	1397/1400	1379/1400	1394/1400	0.9928
7	**1399/1400**	1379/1400	1394/1400	0.9933
8	**1399/1400**	1379/1400	1394/1400	0.9933
9	**1399/1400**	1379/1400	1394/1400	0.9933
10	1397/1400	1379/1400	1394/1400	0.9928

Tabla 2.2: Resultados detallados del **Experimento extendido**, mostrando el rendimiento bajo diferentes transformaciones

Un hallazgo particularmente interesante es que el método demuestra un excelente rendimiento incluso con solo 3 puntos clave, alcanzando un promedio de precisión de 0.9942. Esto sugiere que la técnica es altamente eficiente en términos de recursos computacionales mientras mantiene una robustez excepcional.

2.5.3. Análisis comparativo

Al contrastar los resultados experimentales con otros métodos del estado del arte que utilizan el mismo conjunto de datos (Tabla 2.3), la técnica implementada ha demostrado un rendimiento superior, posicionándose como el método más efectivo hasta el momento, a pesar de no alcanzar un 100 % de reconocimiento.

Autor	Precisión promedio
SCN [Zhang21]	0.7539
BAPmP [Arjun18]	0.8797
(SCF + SCF) (DCA) [Abro19]	0.9196
SA-OAN [Paramarthalingam21b]	0.9434
DSW + Global [Alajlan07]	0.9508
Zernike moment descriptor [Kim00]	0.9588
Esta propuesta	**0.9942**

Tabla 2.3: Comparación con métodos del estado del arte

2.5.4. Análisis detallado del experimento extendido

Los resultados del experimento extendido revelan patrones complejos e interesantes que requieren un análisis profundo para cada tipo de transformación. A continuación, se presenta un análisis detallado de cada aspecto evaluado.

Comportamiento frente al escalamiento

El rendimiento de la técnica frente a escalamiento muestra una evolución notable conforme aumenta el número de puntos de referencia. Se observa una clara mejora progresiva desde los 3-4 puntos iniciales (1388/1400 y 1382/1400 respectivamente) hasta alcanzar un rendimiento óptimo con 7-9 puntos (1399/1400). Esta mejora puede atribuirse a que un mayor número de puntos proporciona redundancia útil para compensar las distorsiones introducidas por el escalamiento.

Es interesante notar que el rendimiento se estabiliza en el rango de 7-9 puntos, pero experimenta una ligera degradación al alcanzar los 10 puntos (1397/1400). Este comportamiento sugiere que existe un punto óptimo en la cantidad de información geométrica necesaria, después del cual la adición de más puntos puede introducir ruido en la representación, afectando negativamente al rendimiento.

Comportamiento frente a la rotación

Las pruebas de rotación exhiben un patrón inverso al observado en el escalamiento. El mejor desempeño se logra con configuraciones más simples de 3-4 puntos (1394/1400),

seguido por una degradación gradual hasta 1379/1400 cuando se utilizan 5 o más puntos. Este fenómeno sugiere que las configuraciones triangulares más simples preservan mejor las relaciones geométricas bajo rotación, mientras que las configuraciones más complejas son más susceptibles a variaciones introducidas por esta transformación.

La eficacia de las representaciones más simples bajo rotación puede explicarse por la naturaleza fundamental de las relaciones geométricas en triángulos básicos, que mantienen mejor sus propiedades invariantes bajo rotaciones. Cuando se aumenta el número de puntos, la complejidad adicional puede introducir inestabilidades en la representación rotada.

Comportamiento frente a la deformación

El aspecto más notable en las pruebas de deformación es la extraordinaria consistencia en el rendimiento. Se mantiene un resultado uniforme de 1394/1400 aciertos independientemente del número de puntos utilizados. Esta estabilidad sugiere una robustez inherente del método frente a deformaciones locales, posiblemente debido a que la representación triangular mantiene efectivamente las relaciones locales incluso bajo deformación.

La consistencia en el rendimiento bajo deformaciones puede atribuirse a dos factores principales: la capacidad de la representación triangular para preservar relaciones locales significativas, y la flexibilidad de la tabla Hash en el proceso de correspondencia de características.

Implicaciones prácticas y optimización

Los resultados del experimento extendido tienen importantes implicaciones para la implementación práctica. Para optimizar el rendimiento, se recomienda utilizar 7-9 puntos cuando se esperen variaciones significativas en escala, 3-4 puntos cuando la rotación sea la principal preocupación, y cualquier número en el rango de 3-10 puntos para casos donde predominen las deformaciones.

La selección del número óptimo de puntos representa un compromiso importante que depende del tipo de transformación esperada en la aplicación específica. Es notable que una configuración simple de 3 puntos ofrece un excelente balance general, alcanzando un promedio de precisión de 0.9942 a través de todas las transformaciones.

2.6. Discusión

Las implicaciones de este trabajo se extienden más allá del campo inmediato de CBIR. En el contexto más amplio de la inteligencia artificial y el aprendizaje profundo, la capacidad de representar y comparar formas de manera eficiente podría tener aplicaciones en áreas como el reconocimiento de objetos, la segmentación de imágenes y el análisis de video en tiempo real. La técnica propuesta también podría servir como base para el desarrollo de sistemas más avanzados que combinen múltiples modalidades de características.

La selección de puntos clave juega un papel importante en la precisión de la técnica propuesta. Se observó que incluso con solo 3 puntos clave, la técnica funciona excepcionalmente bien, lo que sugiere una representación muy eficiente de la forma. Los resultados experimentales obtenidos demuestran la eficacia de la metodología propuesta en la recuperación de imágenes basada en la curvatura de la forma. La precisión alcanzada (0.9942) supera significativamente a los métodos del estado del arte. Es notable la robustez de la técnica propuesta frente a transformaciones como escalamiento, rotación y deformación.

Limitaciones

El análisis de la técnica propuesta revela áreas que requieren mayor investigación. Existe un claro compromiso entre la optimización para diferentes tipos de transformaciones, ya que la configuración óptima para una puede afectar negativamente el rendimiento en otras. Esta observación sugiere la necesidad de desarrollar estrategias adaptativas que puedan ajustar dinámicamente la configuración según el contexto. Además de otras limitaciones detectadas, como:

- La dependencia de contornos bien definidos, lo cual afecta el reconocimiento de objetos en imágenes con bordes difusos o ruido significativo.
- El costo computacional en la fase de pre-procesamiento, el cual podría optimizarse mediante técnicas de paralelización.
- La necesidad de ajustar manualmente ciertos parámetros, como es en el caso de omitir huecos de los objetos, para poder reconocer bien los contornos.

2.6.1. Aplicaciones prácticas

La técnica propuesta encuentra aplicación en diversos campos de la visión computacional y el procesamiento de imágenes, por ejemplo:

- En el sector industrial y manufacturero, la técnica propuesta triunfa por la capacidad para identificar y comparar formas de manera precisa, lo que permite la detección de defectos en productos. La eficiencia computacional de la técnica permite su implementación en sistemas de tiempo real, facilitando la inspección continua en líneas de producción.

- En el campo de la seguridad y vigilancia, la técnica demuestra su utilidad en sistemas de reconocimiento de objetos. La invariancia a rotación y escalamiento permite identificar objetos desde diferentes ángulos y distancias, mientras que la robustez a deformaciones moderadas facilita el seguimiento de objetos en movimiento. Esta característica es especialmente relevante en sistemas de videovigilancia y monitoreo de seguridad.

Estas aplicaciones prácticas demuestran la versatilidad y utilidad de la técnica en escenarios del mundo real, donde la precisión, eficiencia y robustez son requisitos fundamentales. La capacidad de la técnica para adaptarse a diferentes contextos y requerimientos la convierte en una herramienta valiosa para una amplia gama de aplicaciones en visión computacional.

2.7. Conclusiones

La técnica de recuperación de imágenes basada en la curvatura de forma presentada en este capítulo ha demostrado ser efectiva para abordar los objetivos planteados inicialmente. Los resultados experimentales validan que es posible lograr una representación robusta y eficiente de objetos mediante patrones triangulares generados a partir de puntos clave seleccionados estratégicamente.

Las principales contribuciones de este capítulo son:

- El desarrollo de una técnica novedosa para representar y comparar formas que es robusta a transformaciones como rotación, escalamiento y deformaciones moderadas, cumpliendo con el objetivo de crear una técnica invariante a estas transformaciones.

- La implementación de un método eficiente de selección de puntos clave que captura las características esenciales de la forma, permitiendo una representación compacta pero informativa.

- El diseño de una estrategia de indexación basada en triángulos que facilita la búsqueda rápida en grandes colecciones de imágenes, respondiendo al objetivo de desarrollar métodos eficientes de recuperación.

Los experimentos realizados demuestran que la técnica propuesta supera significativamente a los métodos del estado del arte, alcanzando un recall del 99.57 % en el conjunto de datos MPEG-7. Particularmente notable es la robustez de la técnica frente a transformaciones geométricas, manteniendo un recall incluso bajo condiciones de escalamiento (99.9 %), rotación (99.5 %) y corte (99.5 %).

Esta robustez y eficiencia en la representación de formas sienta las bases para el desarrollo de sistemas CBIR más avanzados, que se explorarán en los siguientes capítulos. Sin embargo, también se identificaron áreas de mejora, especialmente en el manejo de objetos con contornos complejos o discontinuos, que servirán como punto de partida para investigaciones futuras.

2.8. Trabajo Futuro

Las direcciones futuras de investigación deberían enfocarse en el desarrollo de métodos de selección adaptativa de puntos, la exploración de representaciones que puedan manejar eficientemente diferentes tipos de transformaciones, y el desarrollo de técnicas de normalización mejoradas que maximicen la robustez del método bajo diversas condiciones, además de:

1. **Optimización de la selección de puntos clave:**

- Desarrollar métodos adaptativos para determinar el número óptimo de puntos clave
- Investigar técnicas para mejorar la robustez en presencia de oclusiones parciales
- Explorar la integración con métodos de aprendizaje automático para la selección de puntos

2. **Mejora del manejo de contornos complejos:**

- Desarrollar técnicas para manejar objetos con múltiples componentes
- Investigar métodos para tratar eficientemente formas con agujeros
- Implementar estrategias para manejar contornos discontinuos

Con estas direcciones de investigación no solo se busca mejorar la técnica propuesta, sino también ampliar su aplicabilidad a diversos escenarios prácticos.

2.9. Comentarios Finales

La técnica propuesta aborda un desafío fundamental en el campo de la visión computacional: la necesidad de representaciones robustas y computacionalmente eficientes para la comparación de formas en imágenes.

La relevancia de este trabajo se hace evidente en el contexto actual del big data y la inteligencia artificial, donde la capacidad de procesar y analizar grandes volúmenes de datos de manera eficiente es muy importante. La técnica desarrollada no solo mejora la eficiencia computacional en la recuperación de imágenes, sino que también establece una nueva tendencia para la representación robusta de formas frente a transformaciones afines.

Los resultados experimentales han demostrado que el enfoque basado en patrones triangulares ofrece ventajas significativas sobre otros métodos. La capacidad de lograr un recall cercano al 100 % para diversas transformaciones geométricas, combinada con una reducción significativa en los recursos computacionales, sugieren que esta técnica podría tener un impacto sustancial en aplicaciones prácticas de visión computacional.

Sin embargo, es importante reconocer que esta técnica representa solo el primer paso en el desarrollo de un sistema CBIR completo y eficiente. Las limitaciones identificadas, particularmente en términos de dependencia de contornos bien definidos y la necesidad de ajuste manual de parámetros, señalan áreas específicas para futuras mejoras. Estas limitaciones serán abordadas en los capítulos subsiguientes de esta tesis doctoral.

El siguiente capítulo marca una transición natural de este trabajo, al introducir códigos de Hadamard como una herramienta para crear representaciones más compactas y eficientes. Esta progresión refleja una evolución desde la representación básica de formas hacia técnicas más sofisticadas que pueden manejar eficientemente grandes volúmenes de datos. La combinación de la robustez de la representación basada en forma con la eficiencia de los códigos de Hadamard promete abrir nuevas posibilidades en el campo de la recuperación de imágenes.

En conclusión, este capítulo ha establecido una base sólida para el desarrollo de sistemas CBIR más eficientes y robustos. Las contribuciones presentadas aquí, junto con las direcciones futuras identificadas, proporcionan un punto de partida prometedor para las innovaciones que se presentarán en los capítulos siguientes.

Capítulo 3

Diseño de una función de pérdida perceptiva con códigos Hadamard

"La innovación distingue a los líderes de los seguidores."

Steve Jobs (1955-2011) y CEO de Apple

Las Redes Neuronales Convolucionales (Convolutional Neural Networks, CNNs) han revolucionado el campo del aprendizaje profundo por su capacidad excepcional para extraer características de alto nivel de las imágenes, por ello son imprescindibles para una amplia gama de aplicaciones de visión computacional.

En este capítulo se propone una técnica para recuperar imágenes basada en su contenido, esta técnica se compone de dos tecnologías que han marcado tendencia en los últimos años, por un lado, se encuentran las redes neuronales convolucionales. Y por otro lado, se encuentran los códigos de Hadamard, conocidos por sus propiedades matemáticas en teoría de la información y procesamiento de señales. Esta propuesta surge como solución a uno de los desafíos más significativos en el campo de la visión computacional: la necesidad de encontrar representaciones de imágenes que sean tanto informativas como computacionalmente eficientes.

Con la técnica propuesta no solo se revoluciona la manera en que se representan y comparan las imágenes, sino que al aprovechar las propiedades de ortogonalidad y equidistancia de los códigos de Hadamard, se logra una codificación que preserva la riqueza informativa de las características profundas tradicionales, mientras se reducen significativamente los requisitos de almacenamiento y procesamiento.

Esta técnica no solo representa un avance teórico en el campo de la recuperación de imágenes, sino que es una solución práctica para el desarrollo de sistemas de visión computacional eficientes y escalables. Sus aplicaciones se extienden más allá del ámbito académico, prometiendo aplicaciones prácticas en sistemas de búsqueda, clasificación de imágenes y otras tareas de visión computacional que requieren procesamiento eficiente de grandes volúmenes de datos.

3.1. Introducción

El aumento de archivos multimedia ha incrementado el uso de vastos repositorios de datos, navegar por dichos repositorios en la búsqueda de información a menudo significa lidiar con datos desorganizados, y problemas de escalabilidad/precisión. Afortunadamente, con el triunfo de las Redes Neuronales Convolucionales Profundas (Deep Neural Network, por sus siglas de inglés) se incentivó una ferviente búsqueda de métodos de comparación de imágenes que permiten recuperar información precisa de una manera rápida y eficiente [Krizhevsky12b].

Las CNNs se destacan en tareas como síntesis y clasificación de imágenes, entre otras aplicaciones como reconocimiento y clasificación de objetos [Donahue14a], sistemas de navegación multimedia [Kratochvíl20], y otras tareas de visión computacional [Azizpour15] por nombrar algunos ejemplos. A pesar de estos avances, uno de los principales desafíos en el campo ha sido la representación efectiva de las clases para la comparación de objetos. Tradicionalmente, la codificación de etiquetas One-hot ha dominado la clasificación en las redes neuronales. Sin embargo, esta representación falla al comparar objetos; los objetos de diferentes clases tienen sin excepción una distancia euclidiana similar (un promedio de $\sqrt{2}$) independientemente de la clase a la que pertenezcan, esto es un gran inconveniente porque

sitúa a todos los objetos a la misma distancia, e imposibilita diferenciarlos entre sí. Como solución a este problema, en las últimas décadas han sido utilizadas las características profundas para este fin. Estas características permiten la comparación de objetos mediante un vector representativo y métricas estándar, como la distancia euclidiana o distancia hamming entre otras [Kloberdanz22], [Kirtas23].

Las características profundas se pueden extraer de la penúltima capa de la red neuronal, su dimensionalidad varía según el modelo neuronal [Pan10]; por ejemplo, VGG utiliza 4096 flotantes, ResNet 2048 y EfficientNet más de 1000. En las redes residuales (ResNet), la capa óptima para la extracción profunda de características sigue siendo ambigua debido a los saltos, incluso si todos los saltos convergen al final.

Navegar en grandes colecciones de datos de imágenes requiere de métodos eficientes de recuperación. Crear índices a imágenes mediante características profundas es una solución que permite navegar en grandes colecciones de una manera efectiva, sin embargo, esta acción exige que los índices sean almacenados en memoria principal, en bases de datos masivas, esta actividad no es viable porque va a exceder los recursos computacionales estándar. En perspectiva, una base de datos de cien millones de imágenes podría consumir alrededor de cuatro terabytes de memoria RAM, esto excede los límites computacionales actuales.

Como alternativa a indexar imágenes, se pueden indexar vectores de características profundas de baja precisión. Aunque esta alternativa es buena opción, exige un proceso de reentrenamiento y cuantificación por parte de los modelos convolucionales, para así poder generar dichas características reducidas [Kloberdanz22], [Kirtas23].

Los recientes avances en el análisis de funciones de pérdida perceptual, documentados por Dosovitskiy et al. [Dosovitskiy16], sugieren que la elección de la función de pérdida puede tener un impacto significativo en la calidad de las representaciones aprendidas. La propuesta de utilizar códigos de Hadamard se alinea con estos hallazgos, ofreciendo un enfoque novedoso para abordar las limitaciones de las funciones de pérdida tradicionales. Al igual que Dosovitskiy, en este capítulo se reconoce la importancia de la función de pérdida en la calidad de las representaciones aprendidas, por tal motivo se propone una función basada en códigos de Hadamard para mejorar la eficiencia de la codificación.

En este capítulo se presenta un método novedoso para crear características pro-

fundas en un espacio significativamente menor: solo 128 bytes por imagen, denominadas Características Profundas de Hadamard (DHF). Para lograrlo, fue reemplazada la salida de la red de codificación one-hot (capa softmax), por una capa densa binaria de 1024 elementos, en la validación se utiliza la distancia de Hamming para encontrar coincidencias con estas características binarias.

En la Figura 3.1 se ilustran las partes fundamentales de la técnica propuesta para obtener las Características Profundas de Hadamard (DHF). Estas partes comprenden la imagen de consulta, el modelo convolucional reentrenado con códigos de Hadamard (que actúa como una caja negra) y la salida que representa la imagen codificada. En esta caja negra es importante resaltar que el modelo convolucional completo está integrado por dos secciones: a) el modelo convolucional base que permanece constante y b) la capa convolucional reentrenada con los códigos de Hadamard.

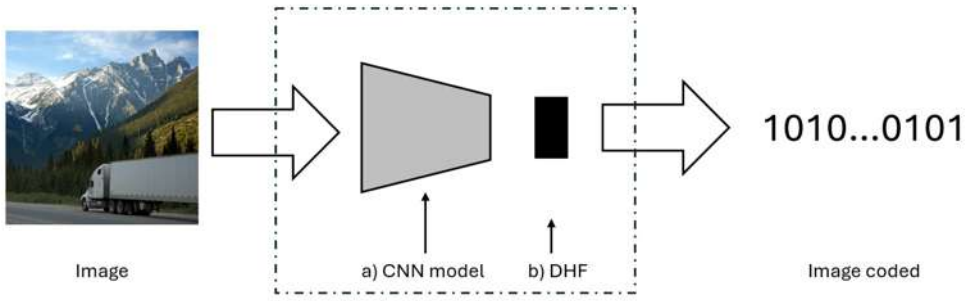


Figura 3.1: Diagrama de la técnica propuesta

Las contribuciones de este capítulo son:

1. Un método innovador con el que se logra una representación compacta y eficiente de características profundas, reduciendo significativamente los requisitos de memoria mientras se preserva la información discriminativa esencial.

2. Un sistema que mantiene la capacidad discriminativa de las características profundas originales en su forma comprimida, garantizando una recuperación efectiva de imágenes semejantes.

3. Una solución efectiva que demuestra robustez frente a diferentes modelos neuronales y tipos de datos, siendo adaptable a diversos escenarios y aplicaciones.

Este capítulo ha sido publicado en Multimedia Tools and Applications bajo el título “Diseño de una función de pérdida perceptual basada en códigos de Hadamard”.

3.2. Trabajo relacionado

Los avances recientes en la generación de características de alto nivel, así como la integración de diferentes modelos neuronales, han revolucionado la recuperación de objetos basados en contenido, porque permiten capturar relaciones contextuales más ricas. Esto ha dado lugar a sistemas más eficientes, robustos y que hacen uso de la transferencia de conocimiento para mejorar el reconocimiento. A continuación se presentan los trabajos más importantes relacionados con estos temas.

3.2.1. Características profundas

Las CNNs han demostrado un rendimiento excepcional en diversas tareas como clasificación y reconocimiento de objetos [Donahue14a] hasta sistemas de navegación web [Kratochvíl20]. En estas tareas se requieren vectores de características para medir la similitud entre objetos. Los vectores de características usualmente son extraídos y/o generados desde modelos convolucionales así como lo muestran Amato et. al. [Amato16]. Estos autores obtienen características profundas de la primera y segunda capa completamente conectadas (fc) del modelo HybridNet (preentrenado en Imagenet). Con estas características, realizan diferentes pruebas respecto a la activación de neuronas, y generan tres características adicionales con el fin de mejorar sus sistemas de reconocimiento:

- ReLU-L2Norm: Estas características son extraídas de la capa fc6, constan de un vector de punto flotante con 4096 elementos, en la comparación, los autores usan distancia euclidiana (L2) o coseno.

- Binario: Estas características corresponde a las características ReLU-L2Norm binarizadas, en la comparación los autores usan la distancia de Hamming.

- Raw: Estas características son similares a las características binarias, pero procedente de la capa fc7 en lugar de fc6.

En los experimentos realizados por Amato et. al. [Amato16], enfatizan la importancia de cada elemento de las características profundas y su aporte en la recuperación de imágenes. En investigaciones posteriores, estos autores desarrollaron un sistema de anotación automática que hereda el uso de sus características propuestas previamente [Amato17b]. Un tiempo después, Amato et al. [Amato17a] diseñaron un sistema interactivo para categorizar fotografías de alimentos compartidas en plataformas sociales utilizando las características profundas de las capas MaxPooling que se encuentran en bloque 5 del modelo GoogLeNet [Aswathy18].

Estos trabajos presentados comparten similitudes con la técnica propuesta, especialmente por la parte de generar representaciones compactas de las características profundas. A diferencia de la técnica propuesta, estos autores extraen características de varias capas previas a la salida de los modelos convolucionales, y en este trabajo se extraen características de la última capa antes de la salida de dichos modelos.

Carrara et. al. [Carrara17] evaluan la resiliencia del modelo Fast OverFeat contra ataques adversarios, empleando características profundas reducidas en dimensión, mediante el algoritmo PCA y características profundas binarias extraídas de la capa pool5. Estos autores extendieron su trabajo original en [Carrara19], donde examinan a fondo el modelo Inception-V3 y amplian sus hallazgos iniciales con el modelo Fast OverFeat, además logran identificar a la capa pool5 como la capa óptima para extraer las características profundas más ricas. Así como en el trabajo de Carrara, en la técnica propuesta se busca reducir la dimensionalidad y uso de memoria de las características profundas, mientras ellos utilizan PCA y la capa pool5, en la propuesta se emplean los códigos de Hadamard para obtener una representación binaria eficiente.

Parola et al. [Parola21] diseñaron e implementaron un sistema de navegación web para recuperar imágenes semejantes a una consulta, usando el modelo convolucional Resnet50. En su sistema, esos autores utilizan las características profundas de la capa conv5_block1_1_conv. Al igual que Parola et al., en la técnica propuesta se tiene la meta

de crear un sistema eficiente para navegar en grandes repositorios de datos en busca de imágenes similares a una consulta. La diferencia en ambos trabajos corresponde al lugar donde son obtenidas las características profundas, y la codificación de Hadamard.

Zhong et al. [Zhong16] realizan un análisis estadístico y exhaustivo de las características profundas como función perceptual, esta actividad revela que tanto la dimensionalidad y la estructura de estas representaciones, tienen un impacto directo en la capacidad del sistema para capturar información semántica relevante. Estos hallazgos respaldan la necesidad de desarrollar representaciones más eficientes y estructuradas de las características profundas. Al igual que Zhong et al., en esta tesis de investigación se realiza un análisis estadístico de las representaciones de características profundas desde diferentes perspectivas.

La técnica propuesta integra conceptos clave de estos trabajos previos, como la eficiencia de las representaciones binarias, la robustez de los códigos de corrección de errores, y el poder de las características profundas, mientras introduce una innovación mediante el uso de códigos de Hadamard como base para la representación de características. Esta integración permite mantener las ventajas de los métodos existentes mientras se mejora significativamente la eficiencia computacional y de memoria.

3.2.2. Códigos Hadamard

El comienzo de la teoría de la información se le atribuye a Claude Shannon por sus destacados avances al evaluar la capacidad de transmitir información, en canales propensos al ruido [Shannon48]. Shannon demostró que era posible transmitir información de manera confiable, incluso en presencia de interferencias, estableciendo los fundamentos matemáticos para el desarrollo de códigos de corrección de errores.

Una alternativa para mitigar los efectos del ruido en la transmisión fue la introducción de redundancia controlada [Shannon48, Proakis08]. Este concepto se materializó de manera brillante a través de los códigos de Hadamard y los códigos de Reed-Solomon [Reed60, Moon20]. Los códigos de Hadamard se destacan por su estructura matemática elegante y sus propiedades de ortogonalidad, que permiten una detección y corrección de errores altamente eficiente. Los códigos de Reed-Solomon complementan esta capacidad con su habilidad para manejar ráfagas de errores y pérdidas de datos [Hanzo11].

La combinación de estos códigos proporciona un sistema robusto de corrección de errores que garantiza una comunicación confiable incluso en condiciones adversas. Su efectividad es tal que, a pesar de las alteraciones de bits inducidas por el ruido, el mensaje original permanece discernible [Hanzo11]. Esta característica ha convertido a los códigos de Hadamard en una herramienta fundamental, no solo para la comunicación digital, sino también para otras aplicaciones donde la integridad y recuperación precisa de la información es muy importante.

A continuación se muestra una exploración más profunda de los códigos de Hadamard, enfocada en su construcción, sus propiedades matemáticas y su aplicación en el contexto de la recuperación de imágenes mediante la construcción de Sylvester [Sylvester67].

La construcción de Sylvester [Sylvester67] es un método recursivo comúnmente utilizado para generar códigos de Hadamard. En este proceso implica la creación de matrices de Hadamard H_{2^r}, donde r determina el número de clases. Para obtener los códigos de Hadamard, se extraen las filas o columnas de estas matrices.

La naturaleza recursiva de la construcción de Sylvester es su característica distintiva, que parte de una matriz base H_2 (Ecuación 3.1), y se construyen matrices más grandes H_{2^r} (Ecuación 3.2) utilizando copias de la matriz del paso anterior. En este proceso iterativo se garantiza que, las matrices resultantes hereden las propiedades fundamentales de los códigos de Hadamard.

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3.1}$$

$$H_{2^r} = \begin{bmatrix} H_{2^{r-1}} & H_{2^{r-1}} \\ H_{2^{r-1}} & -H_{2^{r-1}} \end{bmatrix} \tag{3.2}$$

Ejemplo, para crear una matriz de Hadamard $H_{2^{r=3}=8}$, para 8 clases mediante la construcción de Sylvester, el resultado se puede observar en la Matriz 3.3.

$$H_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} \tag{3.3}$$

Una notable característica de los códigos de Hadamard corresponde la distancia de Hamming entre dos o más códigos de Hadamard (con el mismo valor de r) siempre es igual a $m/2$, donde la cantidad de etiquetas (m) es igual a $m = 2^r$.

A diferencia de la codificación one-hot tradicional, donde todos los objetos de diferentes clases mantienen la misma distancia euclidiana, la técnica propuesta con códigos Hadamard permite una discriminación más fina mediante distancias de Hamming bien definidas y distribuidas uniformemente.

La técnica propuesta se distingue de los métodos tradicionales por no requerir reentrenamiento para cada conjunto de datos. Mientras que otros métodos de codificación sacrifican información semántica por compactabilidad, en la técnica propuesta se mantiene relaciones semánticas gracias a la estructura matemática de los códigos Hadamard, porque estos garantiza una distribución uniforme de distancias entre clases diferentes. Además de esto, la técnica propuesta también se distingue de los métodos anteriores por:

- Su uso innovador de códigos Hadamard como base para la representación de características
- La preservación de relaciones semánticas sin sacrificar eficiencia computacional
- La garantía matemática de distancias uniformes entre clases a través de las propiedades de las matrices de Hadamard
- Obtener una representación compacta y eficiente sin comprometer el recall

Estas diferencias posicionan a la técnica como una solución única y efectiva para el desafío de la recuperación eficiente de imágenes en grandes bases de datos.

3.3. Técnica propuesta

En este capítulo se propone una técnica innovadora, donde se reentrenan modelos convolucionales con códigos de Hadamard para crear representaciones de imágenes compactas y eficientes, esto con el objetivo de mejorar la recuperación de imágenes basada en contenido de grandes bases de datos. Con esta técnica se busca aprovechar las propiedades matemáticas de los códigos de Hadamard para codificar la información de las imágenes de una manera informativa y computacionalmente eficiente.

3.3.1. Fundamento teórico

El fundamento teórico para usar las matrices y códigos de Hadamard en la representación de imágenes se basa en sus propiedades matemáticas, equidistancia y robustez. Hedayat et al. [Hedayat78] han demostrado que las matrices de Hadamard poseen características peculiares para la codificación de información en espacios de alta dimensión, incluyendo la ortogonalidad y la distribución uniforme de distancias. Estas propiedades son valiosas y de gran ayuda en el contexto de la recuperación de imágenes en bases de datos masivas.

La elección de códigos de Hadamard como base para la representación de imágenes se fundamenta en sólidos principios matemáticos y evidencia empírica, acumulada a lo largo de décadas de investigación en teoría de códigos y procesamiento de señales, como son:

1. **Propiedades Matemáticas**. Las matrices de Hadamard poseen características matemáticas que las hacen particularmente adecuadas para la representación de información en espacios de alta dimensión. MacWilliams y Sloane [MacWilliams77] señalan que dichas matrices proporcionan una base ortogonal óptima, con la cual se garantiza una distribución uniforme de distancias entre códigos. Horadam [Horadam12] profundiza en estas propiedades, destacando su relevancia para aplicaciones prácticas en procesamiento de señales.

2. **Robustez y Corrección de Errores**. La estructura matemática de los códigos de Hadamard contribuye significativamente a su robustez. Los trabajos de Shannon [Shannon48] en teoría de la información establecieron las bases para comprender cómo las propiedades de estos códigos permiten una comunicación confiable incluso en presencia de ruido. Hamming [Hamming50] y posteriormente Reed y Solomon [Reed60] expandieron estos conceptos, demostrando la capacidad de corrección de errores inherente a códigos con propiedades similares.

3. **Eficiencia Computacional**. Desde una perspectiva práctica, la naturaleza binaria de los códigos de Hadamard facilita implementaciones computacionalmente eficientes. Hanzo et al. [Hanzo11] han demostrado cómo estas propiedades pueden aprovecharse en sistemas modernos de procesamiento de información, proporcionando un equilibrio óptimo entre eficiencia y robustez.

4. **Garantías Teóricas**. La teoría de la información establecida por Shannon [Shannon48], proporciona el marco teórico que sustenta la eficacia de los códigos de Hadamard. Sus propiedades de ortogonalidad y distancia uniforme garantizan una separación óptima entre diferentes representaciones, lo cual es crucial para tareas de clasificación y recuperación.

Esta fundamentación teórica sugiere a los códigos de Hadamard como una elección robusta para lograr un equilibrio entre eficiencia computacional y precisión en la recuperación de imagenes semejantes.

3.3.2. Ventajas de la técnica propuesta

Lo más importante de esta propuesta gira en torno al concepto de utilizar códigos de Hadamard como una representación compacta y robusta, para el etiquetado de clases en redes neuronales profundas. Este enfoque ha demostrado una excelente eficiencia en resultados experimentales. La mayor ventaja es que ofrece una combinación excepcional de eficiencia y eficacia, especialmente cuando se yuxtapone a métodos tradicionales como la reducción de precisión.

La reducción de precisión cuando se lleva al límite aminora considerablemente el consumo de memoria principal a 8 bits, 4 bits, incluso a un 1 bit. Sin embargo, este hecho trae consigo inestabilidades numéricas. Estas limitaciones se vuelven más pronunciadas a medida que se reduce aún más la precisión (en la sección de resultado se muestra este efecto). En los experimentos se hicieron diferentes pruebas para reducir la precisión desde 8 hasta 1 bit. Sin embargo, en términos de eficiencia, los DHF superan notablemente el enfoque de precisión reducida.

La flexibilidad que ofrecen las DHF se encuentra en gran medida a su etiquetado basado en las matrices de Hadamard. Como estas matrices tienen propiedades matemáticas independientes de los modelos convolucional, esto ofrece flexibilidad en su aplicación en varias modelos neuronales. Otras ventajas de este método se pueden resumir de la siguiente manera:

- **Compactabilidad**: La representación de las DHF es concisa, incluso más que las estrategias de reducción de precisión más agresivas.
- **Robustez**: A diferencia de la reducción de precisión, que puede introducir inestabilidades numéricas, el método aquí propuesto conserva robustez, haciéndolo menos susceptible a perturbaciones adversas.
- **Eficiencia**: Los resultados experimentales muestran una precisión superior a diferencia de modelos con precisión reducida.
- **Uso de memoria**: Más allá de la compactabilidad de la representación, los DHF conducen a importantes ahorros de memoria para su implementación en dispositivos con recursos limitados.

3.4. Implementación de la propuesta

Las características profundas de Hadamard (Deep Hadamard Features, DHF por sus siglás del inglés) se originan en la capa de salida de un modelo convolucional reentrenado con los códigos Hadamard como alternativa a one-hot. La implementación de la técnica propuesta se divide en dos etapas principales como se muestra en la Fig. 3.2

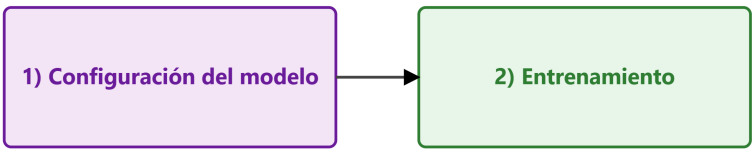


Figura 3.2: Etapas de la técnica propuesta

1. **Configuración del modelo**: Aquí, la salida softmax de una CNN previamente entrenada se reemplaza con una capa densa, cuya longitud corresponde a la longitud del código Hadamard. Todas las capas del modelo CNN previamente entrenado se mantienen estáticas, excepto la última capa densa recién introducida. Además, se asignan etiquetas para relacionar las clases que reconoce el modelo CNN con los códigos Hadamard. Estas etiquetas se generan usando el método de Sylvester con entradas $\{-s, s\}$, donde la elección canónica $s = 1$ y produce el rango $\{-1, 1\}$.

2. **Entrenamiento**: Al inicio de esta fase se designa un código Hadamard único a cada clase dentro del conjunto de datos. Similar al ejemplo para solo 10 clases que se muestra en la Tabla 3.1, en esta tabla se muestran una descripción del número de clases en conjunto con sus codificaciones one-hot y Hadamard.

   Los vectores de Sylvester con entradas $\{-s, s\}$ funcionan como prototipos a lo largo de esta fase de entrenamiento. En esta fase se ajusta el tamaño de lote desde 4 hasta un máximo de 64 imágenes. La métrica del error cuadrático medio (MSE) fue adoptada para evaluar la discrepancia entre los códigos de Hadamard durante el entrenamiento como función de pérdida. Durante la validación (después de la fase de entrenamiento) los vectores se binarizan y la distancia de Hamming reemplaza a MSE para las evaluaciones.

Durante la fase de entrenamiento, la función de pérdida perceptual aprovecha una base de datos completa de imágenes etiquetadas. Una vez entrenado el modelo CNN resultante, está función sirve para comparar imágenes de diferentes bases de datos. Es importante mencionar que la cantidad de imágenes parar entrenar un modelo CNN impacta directamente en la eficiencia, y la cantidad de parámetros de dicho modelo influye en el tiempo de evaluación.

Tabla 3.1: Ejemplo de codificación one-hot y Hadamard para un conjunto de datos con diez clases.

clase	Codificación one-hot	Codificación de Hadamard
0	1 0 0 0 0 0 0 0 0 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1	0 1 0 0 0 0 0 0 0 0	1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
2	0 0 1 0 0 0 0 0 0 0	1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0
3	0 0 0 1 0 0 0 0 0 0	1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1
4	0 0 0 0 1 0 0 0 0 0	1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0
5	0 0 0 0 0 1 0 0 0 0	1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 1
6	0 0 0 0 0 0 1 0 0 0	1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1
7	0 0 0 0 0 0 0 1 0 0	1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0
8	0 0 0 0 0 0 0 0 1 0	1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
9	0 0 0 0 0 0 0 0 0 1	1 0 1 0 1 0 1 0 0 1 0 1 0 1 0 1

La clasificación en este trabajo sigue un enfoque diferente al método tradicional de codificación one-hot. Mientras que en la codificación one-hot la clase se determina por la neurona con la mayor activación de salida, en la técnica propuesta la clasificación se realiza calculando la distancia de Hamming entre el vector de salida y todos los códigos de Hadamard disponibles. La clase asignada corresponde a aquella cuyo código de Hadamard presente la menor distancia de Hamming con respecto al vector de salida.

Aprovechando las observaciones empíricas de Hoyos et. al. [Hoyos21] acerca de la distancia entre los códigos de hadamard, estos autores concluyeron que la distancia entre clases puede aumentar más allá de $m/2$, cambiando los valores canónicos de los vectores de Sylvester ($\{-1, 1\}$). Hoyos et. al. encontraron que los valores canónicos de $\{-25, 25\}$ aumentan la robustez del modelo CNN contra ataques adversarios. Sin embargo, como los ataques adversarios no son tema de este trabajo de tesis, el valor canónico utilizado corresponde a $\{-1, 1\}$.

En el Algoritmo 2 se muestran el proceso para entrenar modelos convolucionales con los códigos de hadamard. Este comienza determinando el tamaño necesario de los códigos (línea 1), seguido por la generación de la matriz Hadamard usando el método de construcción de Sylvester (línea 2). El proceso de adaptación del modelo convolucional comienza con la clonación del modelo CNN preentrenado (línea 3), reemplazando su capa softmax original por una nueva capa densa cuyo tamaño corresponde a la dimensión de

los códigos Hadamard, y congelando todas las capas excepto la última para preservar el conocimiento preentrenado (línea 4). La etapa de entrenamiento utiliza lotes de 4 hasta 64 imágenes (línea 5) y se estructura en épocas (líneas 6-14), donde cada iteración procesa las imágenes del lote, obtiene sus correspondientes códigos Hadamard, calcula la pérdida mediante el error cuadrático medio (MSE) (línea 8), y actualiza los pesos de la última capa mediante retropropagación (línea 9). Durante la validación, se emplea la distancia Hamming para evaluar el rendimiento, se permite una terminación temprana si se alcanza la convergencia a las pocas épocas de haber iniciado (líneas 11-13).

Algoritmo 2 Generación de Características Profundas de Hadamard

Entrada: Modelo CNN pre-entrenado M, Base de datos de imágenes D

Salida: Modelo reentrenado M' con características DHF

1: $r \leftarrow$ Determinar tamaño de código
2: $H \leftarrow$ GenerarMatrizHadamard(2^r)
3: $M' \leftarrow$ Reemplazar capa softmax de M por una capa densa
4: Congelar todas las capas de M' excepto la última
5: $batch\_size \leftarrow [4 - 64]$ {Tamaño de lote}
6: **para** cada época **hacer**
7: **para** cada lote B en D **hacer**
8: Calcular pérdida MSE entre predicciones con H
9: $M' \leftarrow$ Actualizar pesos de última capa vía backpropagation
10: **fin para**
11: **si** criterio de convergencia alcanzado **entonces**
12: **break**
13: **fin si**
14: **fin para**
15: **devolver** Modelo M' entrenado con códigos de hadamard

3.4.1. Criterios de Implementación

En la técnica propuesta se utilizan matrices H_{2^r} binarias (Q_{2^r}), similar a la Matriz 3.4.

$$Q_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{3.4}$$

Esta matriz binaria Q, se utiliza en la propuesta durante la etapa de validación con el fin de ahorrar memoria. Esta idea tiene su fundamento en el peso de las neuronas, por lo general es mayor a cero, al igual, la mayoría de las funciones de activación regresan valores positivos, como es el caso de la función ReLU (Rectified Linear Unit) y sus variantes. Con esto en mente y en aras de reducir el uso de memoria principal, la matriz H_{2^r} es binarizada a razón de reemplazar los -1's por 0's, esta acción da lugar a la matriz binaria Q.

La relación entre el tamaño de la matriz de hadamard con el número de clases utilizadas en el problema a resolver, basta con 1000 para el etiquetado de la base de datos utilizada, por lo tanto, es suficiente un $k = 10$ debido a qué $2^{k=10} = 1024$.

3.4.2. Bases de datos y modelos utilizados

El campo del aprendizaje profundo está marcado por una innovación incesante, con modelos neuronales más nuevos que traspasan continuamente los límites de lo posible. Además de los modelos fundamentales, el caso de estudio que se presenta en este capítulo profundiza en algunos de los últimos pioneros en las CNN, cada uno de estos introduce técnicas y paradigmas novedosos que no podían quedar exentos a un análisis profundo. Los siguientes modelos neuronales fueron reentrenados en los experimentos, para mapear cada imagen del conjunto de entrenamiento a un código de hadamard:

VGGNet: Finalista en la competencia ImageNet 2014, VGGNet tiene variantes como VGG16 y VGG19 que se distinguen por su número de capas. Favorable para la similitud de imágenes perceptuales, el desafío de VGGNet radica en sus considerables 138 millones de parámetros.

ResNet: ResNet, ganador de ImageNet 2015, este modelo introduce conexiones de salto que imitan las conexiones de neuronas distantes en redes biológicas, con versiones desde ResNet19 a ResNet152, las variantes populares son ResNet50 y ResNet101. El principal desafío con ResNet es la dificultad para extraer características profundas debido a estos saltos.

EfficientNet:. EfficientNet mejora la eficiencia al limitar los parámetros y FLOPS de los modelos CNN. EfficientNet esta compuesto para escalar uniformemente la profundidad, el ancho y la resolución de las imágenes, su innovación radica en un diseño compatible con dispositivos móviles, junto con una estrategia de escalado para lograr una precisión óptima.

MNASNet1-3 [Tan19a]: MNASNet es un modelo neuronal óptimo, ligero y eficaz que aprovecha el aprendizaje por refuerzo. Este modelo está diseñado para entornos móviles, garantiza la eficiencia energética sin comprometer la precisión.

ConvNeXt-Large [Liu22b]: ConvNeXt-Large infunde los principios de convolución de grupo en las CNN. Este modelo al agrupar canales logra reducir la sobrecarga computacional sin perder calidad de representación. Este modelo muestra el potencial de las convoluciones grupales en tareas de clasificación de imágenes a gran escala.

ViT-H-14 [Schuhmann22]: Vision Transformer (ViT) utiliza transformes para tareas de procesamiento del lenguaje natural, y clasificación de imágenes. La variante ViT- H- 14 representa un modelo más grande de la familia ViT, el cual aprovecha transformadores más potentes y precisos. Al tratar una imagen como una secuencia de parches, ViT captura dependencias de largo alcance y patrones intrincados que podrían eludir las CNN tradicionales.

RegNet-Y-128GF [Radosavovic20]: RegNet es único en su enfoque del diseño neuronal. RegNet emplea un diseño simple que permite la generación de una gran cantidad de modelos neuronales. La variante Y-128GF representa una configuración específica en

términos de ancho, profundidad y resolución.

MaxVit-T [Tu22]: Un modelo neuronal híbrido que combina las fortalezas de las CNN con transformadores. MaxVit-T integra información espacial de las CNN en los transformadores, asegurando que el modelo capture características de imagen tanto locales como globales. Esta fusión ofrece mejores resultados en tareas de clasificación de imágenes a gran escala.

Swin-V2-B [Liu22a]: Un sucesor del Swin Transformer original, Swin-V2-B está optimizado tanto para eficacia como para eficiencia. Este modelo neuronal introduce ventanas deslizantes y fusión de tokens jerárquicos para capturar un contexto más amplio en las imágenes.

En la Tabla 3.2 se muestra un resumen de los modelos convolucionales utilizados en la implementación de la técnica propuesta, así como la cantidad de parámetros y sus características principales.

Tabla 3.2: Resumen de las características principales de los modelos CNN utilizados

Modelo	**Año**	**Parámetros**	**Característica principal**
VGGNet	2014	138M	Simplicidad y eficacia
ResNet	2015	25.6M - 60.2M	Conexiones residuales
EfficientNet	2019	5.3M - 66M	Escalado uniforme
MNASNet1-3	2019	5.3M	Optimizado para móviles
ConvNeXt-Large	2022	197.8M	Convoluciones grupales
ViT-H-14	2020	633.5M	Transformers para visión
RegNet-Y-128GF	2020	644.8M	Diseño sistemático
MaxVit-T	2022	30.9M	Híbrido CNN-Transformer
Swin-V2-B	2022	87.9M	Ventanas desplazadas

Los modelos presentados en esta sección son ampliamente utilizados en la recuperación de imágenes basada en contenido, un ejemplo de ello son los modelos ViT y ConvNeXt-Large, estos modelos forman parte de CLIP, una de las aplicaciones más potentes de la actualidad en cuanto a la recuperación de imágenes y descripción de la escena.

3.5. Resultados experimentales

En esta sección se detallan los experimentos hechos con códigos Hadamard para el etiquetado de clases, comparando su eficacia con enfoques alternativos. Estas alternativas incluyen el uso de características profundas, la aplicación de una transformación lineal W a la capa de salida del modelo final, y la reducción de precisión desde 32 bits a 16, 8 y 4 bits (media precisión, un cuarto de precisión y medio byte respectivamente). Para garantizar la robustez estadística de los resultados, cada experimento se repitió al menos 10 veces desde su entrenamiento por cada modelo convolucional.

Los modelos utilizados pueden ser descargados desde el sitio oficial de PyTorch. Estos modelos se encuentran previamente entrenados en ImageNet: VGG16 [Simonyan15], Resnet50 [He16a], Resnet101 [He16a], EfficientNet b0-b3 citeEfficientNet, MNASNet1-3 [Tan19a], ConvNeXt-Large [Liu22b], ViT-H-14 [Schuhmann22], MaxVit-T [Tu22], RegNet-Y-128GF [Radosavovic20], y Swin-V2-B [Liu22a], se incorporaron los clasificadores kNN (k Vecino más cercano) y HSP (Half Space Proximal) [Talamantes22], como clasificadores basados en instancias.

Las bases de datos Cifar-100, Mini-Imagenet, Coco 2017 e ImageNet son puntos de referencia estándar para la clasificación de imágenes, estas bases fueron utilizadas para evaluar las HDF. La tabla 3.3 ofrece una descripción general y concisa de cada base de datos. Para una evaluación comparativa, estos conjuntos de datos se clasificaron según la dificultad de clasificación, teniendo en cuenta la resolución y la cantidad total de imágenes.

Tabla 3.3: Descripción de las bases de datos.

Base de datos	Muestras	Resolución	Clases
Cifar-100 [Krizhevsky09]	60 k	32x32	100
Mini-Imagenet [Vinyals16]	64 k	80x80	64
Coco [Lin14]	200 k	224x224	80
Imagenet [Deng09b]	1.3 M	224x224	1000

Parámetros de los modelos convolucionales

La tabla 3.4 ilustra los requisitos de memoria tanto para los vectores de características profundas como para los vectores de características profundas de Hadamard en varios modelos CNN. En particular, el enfoque de este trabajo exige solo 1024 bits (o 128 bytes) independientemente del modelo, las columnas de la tabla enumeran el nombre del modelo CNN, el número total de parámetros, las dimensiones y requisitos de bits para las características respectivas.

Tabla 3.4: Parámetros por modelo, dimensionalidad y número de bits necesarios por vector de características.

Modelo	Parametros	Características profunas		HDF	
		Dim	Bits	Dim	Bits
VGG16	138 M	4096	131072		
ResNet101	45 M	2048	65536		
ResNet50	26 M	2048	65536		
Efficientnet b3	12 M	1536	49152		
Efficientnet b2	9.2 M	1408	45056		
Efficientnet b1	7.8 M	1280	40960		
Efficientnet b0	5.3 M	1280	40960	1024	1024
MNASNet1-3	6.3M	1280	40960		
ConvNeXt-Large	197.8M	1536	49152		
ViT-H-14	633.5M	1280	40960		
RegNet-Y-128GF	644.8M	7392	236544		
MaxVit-T	30.9M	512	16384		
Swin-V2-B	87.9M	1024	32768		

La tabla 3.5 contrasta los modelos entrenados de PyTorch con los modelos reentrenadas con códigos de Hadamard en ImageNet. Es digno de mención que modelo Resnet101 reentrenado en este trabajo supera a la versión estándar de PyTorch Resnet101.

Tabla 3.5: Datos de referencia de los modelos de PyTorch vs modelos de Hadamard reentrenados.

Modelo	Baseline		Hadamard reentrenado	
	@1	@5	@1	@5
VGG16	71.592	90.382	69.97	82.12
ResNet101	77.374	93.546	71.31	**94.41**
ResNet50	76.130	92.862	66.03	90.74
Efficientnet b3	82.008	96.054	71.312	90.718
Efficientnet b2	80.608	95.310	72.716	91.604
Efficientnet b1	78.642	94.186	70.492	90.516
Efficientnet b0	77.692	93.532	65.676	88.61
MNASNet1 3	76.506	93.522	73.122	88.012
ConvNeXt-Large	84.414	96.976	82.650	90.781
ViT-H-14	88.552	98.694	84.896	90.862
RegNet-Y-128GF	88.228	98.682	80.010	98.556
MaxVit-T	83.700	96.722	80.762	90.786
Swin-V2-B	84.112	96.864	78.267	95.886

3.5.1. Tarea de clasificación

El objetivo de la clasificación es evaluar el potencial de aprendizaje por transferencia de conocimiento de los vectores de Hadamard. La idea general es entrenar una red en un dominio o base de datos y luego usarla en un dominio o base de datos diferente sin ningún entrenamiento adicional. Para este fin, los clasificadores basados en instancias sirven como un mecanismo de prueba ideal, aquí es donde el muy conocido algoritmo k-vecino más cercano (kNN) toma ventaja. Sin embargo, el clasificador HSP (Half Space Proximal) recientemente introducido (o HSP(voto)) presenta una alternativa que no depende de especificar parámetros a diferencia de kNN, el cual requiere un número k de vecinos [Talamantes22].

El objetivo de la clasificación es evaluar el potencial de transferencia de conocimiento con los códigos de Hadamard. Esta técnica permite entrenar una red en un dominio o base de datos específica y posteriormente utilizarla en un dominio diferente sin necesidad de entrenamiento adicional. Para evaluar este proceso, los clasificadores basados en instancias sirven como mecanismo de prueba ideal, siendo el algoritmo k-vecino más cercano (kNN) una opción ampliamente utilizada. Sin embargo, el clasificador HSP (Half Space Proximal) recientemente introducido (o HSP(voto)) ofrecen una alternativa que, a diferencia de kNN,

no requiere especificar parámetros como el número de vecinos [Talamantes22].

La base del clasificador HSP es el grafo HSP presentado por Chávez et. al. [Chavez06]. Este clasificador HSP utiliza el grafo HSP para identificar los vecinos más cercanos de un objeto de consulta y emplea hiperplanos para separar los objetos en vecindades como se muestra en la Fig. 3.3 [Chavez06].

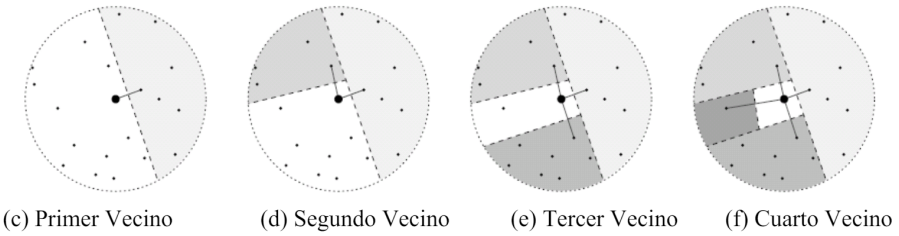


Figura 3.3: Clasificador HSP [Chavez06].

En la implementación de la técnica propuesta son adoptadas dos estrategias para mejorar la eficiencia del clasificador HSP. En primer lugar, usar solo una fracción (entre el 1 % y el 5 %) de los vecinos más cercanos de la base de datos. En segundo lugar, emplear el índice HNSW [Malkov20a] ($k = 250$) a través de HNSWLIB. Las medidas de distancia son; la distancia euclidiana a las características profundas, y la distancia de Hamming a las características profundas de hadamard.

3.5.2. Transferencia de conocimiento

La premisa fundamental de este trabajo postula que un modelo CNN entrenado con una base de datos grande como ImageNet, se puede implementar sin problemas en una base de datos completamente diferente sin necesidad de un reentrenamiento adicional, lo que garantiza tasas de recuperación consistentes y sólidas en todo momento. Un elemento central de este esfuerzo es la capacidad de navegar eficientemente en una base de datos multimedia. Lograr tasas de clasificación altas a través de kNN, sirve como testimonio de la eficacia de la representación aprendida, lo que resalta aún más su potencial como función de pérdida de percepción.

La Tabla 3.6 y 3.8 muestran el recall obtenido para kNN y HSP respectivamente.

Este notable aumento es visualmente evidente en estas tablas, donde las tasas de recuperación se destacan por su notable incremento. Mientras que el primero emplea modelos estándar preentrenados de PyTorch para la extracción de características, el segundo involucra los modelos reentrenados con códigos de Hadamard. Los resultados experimentales obtenidos son comparables en términos de clasificación, estos muestran un contraste en el consumo de memoria. Por ejemplo, mientras que VGG16 exige 16 KB por imagen para sus funciones profundas, Hadamard solo exige 128 bytes.

Tabla 3.6: Resultados con k-NN

		Características profundas						**Hadamard**					
Modelo	Base de datos	kNN			kNN(Voto)			kNN			kNN(Voto)		
		@1	@5	@10	5	7	9	@1	@5	@10	5	7	9
1	I	.352	.596	.703	.460	.473	.482	.369	.620	.725	.496	.484	.474
	II	.683	.868	.917	.768	.772	.774	.753	.887	.921	.810	.808	.806
	III	.546	.767	.833	.651	.646	.639	.503	.726	.799	.612	.609	.604
	IV	.612	.808	.853	.690	.690	.692	.611	.802	.844	.685	.686	.687
2	I	.490	.725	.814	.590	.600	.601	.503	.725	.807	.599	.590	.583
	II	.832	.940	.961	.876	.878	.879	.844	.929	.950	.872	.868	.870
	III	.602	.801	.858	.682	.675	.671	.548	.753	.819	.641	.634	.630
	IV	.628	.830	.876	.719	.721	.720	.625	.798	.831	.687	.689	.691
3	I	.458	.707	.795	.559	.563	.573	.476	.704	.792	.580	.570	.564
	II	.790	.921	.947	.852	.853	.855	.814	.921	.944	.856	.852	.849
	III	.624	.823	.872	.703	.698	.696	.558	.760	.823	.645	.640	.634
	IV	.649	.844	.884	.732	.734	.733	.639	.802	.832	.697	.698	.701
4	I	.569	.779	.848	.640	.650	.660	.488	.711	.793	.582	.581	.574
	II	.858	.932	.950	.885	.881	.880	.814	.914	.939	.855	.849	.852
	III	.624	.777	.839	.660	.660	.659	.565	.771	.835	.652	.648	.647
	IV	.618	.774	.800	.680	.681	.681	.667	.797	.830	.708	.705	.704
5	I	.542	.766	.835	.618	.627	.634	.480	.715	.798	.591	.581	.571
	II	.844	.931	.951	.870	.871	.874	.834	.925	.949	.865	.863	.864
	III	.617	.777	.834	.661	.656	.653	.523	.692	.799	.611	.610	.603
	IV	.636	.802	.832	.699	.702	.703	.660	.782	.811	.699	.697	.695
6	I	.541	.757	.833	.611	.618	.627	.522	.750	.828	.622	.611	.602
	II	.849	.932	.952	.865	.867	.873	.834	.931	.953	.873	.872	.869
	III	.629	.788	.846	672	.667	.662	.577	.782	.848	.666	.663	.660
	IV	.625	.789	.820	.686	.689	.690	.669	.785	.814	.708	.706	.705
7	I	.515	.739	.818	.593	.600	.610	.459	.693	.779	.566	.558	.549
	II	.837	.933	.955	.861	.862	.868	.813	.916	.942	.852	.850	.847
	III	.602	.783	.847	.666	.663	.661	.546	.763	.830	.646	.639	.636
	IV	.616	.800	.837	.694	.694	.693	.652	.777	.805	.695	.693	.693
8	I	.410	.593	.702	.460	.472	.480	.389	.633	.728	.501	.500	.498
	II	.781	.870	.903	.801	.802	.801	.837	.880	.974	.843	.844	.844
	III	.646	.776	.846	.753	.748	.732	.515	.759	.812	.634	.631	.628
	IV	.752	.840	.859	.763	.783	.782	.701	.805	.961	.735	.736	.736

En esta Tabla 3.6 y 3.8, el modelo 1 es VGG16, el modelo 2 es ResNet101, el modelo 3 es ResNet, el modelo 4 es Efficientnet b3, el modelo 5 es Efficientnet b2, el modelo 6 es Efficientnet b1, el Modelo 7 es Efficientnet b0 y el Modelo 8 es Swin-V2-B. El conjunto de datos I es Cifar-100, el conjunto de datos II es Mini-Imagenet, el conjunto de datos III es Coco y el conjunto de datos IV es Imagenet.

Tabla 3.7: Diferencias entre características (DF - Hadamard) de la Tabla 3.6.

Modelo	Conjunto de datos	kNN			HSP	
		@1	@5	@10	HSP(Vote)	HSP+
1	I	-.017	-.024	-.022	-.036	-.089
	II	-.070	-.019	-.004	-.042	-.122
	III	+.043	+.041	+.034	+.039	-.097
	IV	+.001	+.006	+.009	+.005	+.004
2	I	-.013	+.000	+.007	+.009	-.052
	II	-.012	+.011	+.011	+.004	-.010
	III	+.054	+.048	+.039	+.043	-.075
	IV	+.003	+.032	+.045	+.032	+.060
3	I	+.018	+.003	+.003	-.008	-.068
	II	-.024	+.000	+.003	-.004	-.018
	III	+.066	+.063	+.049	+.058	-.043
	IV	+.010	+.042	+.052	+.038	+.062
4	I	-.081	-.068	-.055	-.058	-.039
	II	-.044	-.018	-.011	-.034	-.039
	III	-.059	-.006	-.004	-.008	-.115
	IV	+.049	+.023	+.030	+.028	+.120
5	I	-.062	-.051	-.037	-.027	-.073
	II	-.010	-.006	-.002	-.005	-.032
	III	-.094	-.085	-.035	-.039	-.004
	IV	+.024	-.020	-.021	-.009	+.059
6	I	-.019	-.007	-.005	-.005	-.041
	II	-.015	-.001	+.001	+.008	+.041
	III	-.052	-.006	+.002	-.002	+.105
	IV	+.044	-.004	-.006	+.027	+.091
7	I	-.056	-.046	-.039	-.027	+.027
	II	-.024	-.017	-.013	-.009	+.024
	III	-.056	-.020	-.017	-.020	+.077
	IV	+.036	-.023	-.032	+.008	+.046
8	I	-.021	+.040	+.026	+.041	+.067
	II	+.056	+.010	+.071	+.042	+.016
	III	-.131	-.017	-.034	-.104	+.055
	IV	-.051	-.035	+.102	-.028	-.063

En las Tablas 3.7 y 3.9 se presenta un análisis comparativo entre las características profundas y Hadamard, mostrando las diferencias de rendimiento en los escenarios de k-NN y HSP respectivamente. Los valores negativos (en rojo) indican casos donde las características profundas superaron a Hadamard, mientras que los valores positivos (en azul) señalan lo contrario.

Tabla 3.8: Resultados con HSP

		Características profundas			Hadamard		
Modelo	Conjunto de datos	Hsp(Voto)	Hsp+	Avg. # neighbors	Hsp(Voto)	Hsp+	Avg. # Neighbors
1	I	.503	.777	15	.495	.866	26
	II	.806	.951	13	.812	.973	18
	III	.668	.767	7	.614	.864	15
	IV	.712	.841	8	.703	.837	8
2	I	.622	.887	17	.593	.939	36
	II	.894	.979	12	.878	.989	18
	III	.718	.821	8	.639	.896	20
	IV	.737	.863	9	.702	.803	6
3	I	.596	.869	17	.572	.937	39
	II	.871	.972	12	.850	.990	23
	III	.737	.840	8	.652	.883	16
	IV	.749	.867	8	.711	.805	6
4	I	.678	.886	14	.588	.925	31
	II	.898	.944	7	.851	.983	18
	III	.680	.780	6	.665	.895	16
	IV	.685	.740	5	.722	.860	17
5	I	.652	.888	17	.587	.917	27
	II	.896	.955	8	.871	.987	16
	III	.680	.798	7	.622	.802	16
	IV	.711	.781	5	.708	.840	17
6	I	.644	.893	17	.616	.934	28
	II	.890	.945	8	.875	.986	17
	III	.687	.808	7	.670	.913	18
	IV	.691	.748	5	.718	.839	16
7	I	.633	.884	17	.558	.911	28
	II	.891	.960	9	.857	.984	17
	III	.680	.815	8	.640	.892	17
	IV	.702	.782	6	.710	.828	15
8	I	.598	.873	17	.590	.940	37
	II	.877	.975	10	.849	.991	20
	III	.740	.840	8	.649	.895	18
	IV	.749	.868	8	.712	.805	6

Tabla 3.9: Diferencias entre características (DF - Hadamard) para la Tabla 3.8.

Modelo	Conjunto de datos	HSP(Voto)	HSP+	Vecinos
1	I	-.008	+.089	+11
	II	+.006	+.022	+5
	III	-.054	+.097	+8
	IV	-.009	-.004	0
2	I	-.029	+.052	+19
	II	-.016	+.010	+6
	III	-.079	+.075	+12
	IV	-.035	-.060	-3
3	I	-.024	+.068	+22
	II	-.021	+.018	+11
	III	-.085	+.043	+8
	IV	-.038	-.062	-2
4	I	-.090	+.039	+17
	II	-.047	+.039	+11
	III	-.015	+.115	+10
	IV	+.037	+.120	+12
5	I	-.065	+.029	+10
	II	-.025	+.032	+8
	III	-.058	+.004	+9
	IV	-.003	+.059	+12
6	I	-.028	+.041	+11
	II	-.015	-.041	+9
	III	-.017	+.105	+11
	IV	+.027	+.091	+11
7	I	-.075	-.027	+11
	II	-.034	-.024	+8
	III	-.040	+.077	+9
	IV	+.008	+.046	+9
8	I	-.008	-.067	+20
	II	-.028	-.016	+10
	III	-.091	-.055	+10
	IV	-.037	+.063	-2

3.5.3. Transformación lineal versus reentrenamiento

El reentrenamiento es un proceso efectivo de reajuste, sin embargo, esto conlleva un conjunto de desafíos que exigen destreza computacional, que consumen tiempo de procesamiento y requieren un ajuste meticuloso para garantizar que el modelo reentrenado conserve las características esenciales del original. El objetivo, por lo tanto, es encontrar un término medio: un método que encapsule la esencia de las características profundas de Hadamard sin la necesidad de entrenamiento.

La transformación lineal, denotada por W, fue diseñada para transformar códigos One-Hot estándar en códigos Hadamard directamente. La hipótesis fundamental planteaba que esta transformación, al aplicarse tanto a vectores codificados de manera directa como a vectores de características profundas, podría generar vectores con propiedades similares a las características DHF. Sin embargo, como se detalla en los experimentos, la transformada W no estuvo a la altura de las expectativas iniciales como se muestra en la Tabla 3.10. En esta tabla las filas de *Características Profundas* muestran las tasas de recuperación base de los modelos convolucionales, y la fila de *Hadamard lineal 1024* y *Hadamard lineal 4096* corresponden a las tasas de recuperación después de usar la transformación W con características profundas de la capa final (la codificación One-Hot) y la penúltima capa (características profundas), respectivamente.

Los resultados experimentales obtenidos con *Hadamard lineal 1024* y *Hadamard lineal 4096* son muy pobres, una explicación a este efecto podría relacionarse al proceso de aprendizaje del modelo CNN. Es concebible que el conocimiento acumulado por los modelos durante su fase de aprendizaje sea multifacético y no pueda emularse mediante una simple transformación de códigos One-hot a códigos Hadamard. Además, la no linealidad introducida por la función de activación ReLU también podría desempeñar un papel crucial en esta divergencia. El impacto potencial de la función de activación subraya la complejidad de las operaciones de las redes neuronales y cómo es posible que no siempre se alineen con las transformaciones lineales.

A la luz de estos hallazgos, si bien un proceso de transformación simplificado es innegablemente atractivo, resulta evidente que la profundidad y complejidad del conocimiento

incorporado en los modelos entrenados, no se puede replicar o transformar fácilmente sin comprometer el desempeño.

3.5.4. Reducción de precisión

Los recursos computacionales son limitados para cualquier dispositivo actual. Reducir la precisión de bits de los pesos y activaciones de las redes neuronales es una vía convincente para disminuir las demandas computacionales y de memoria de los modelos profundos. Sin embargo, las reducciones traen consigo una serie de ventajas y desventajas asociadas.

En este trabajo se redujo la cantidad de bits a varios niveles: 16 bits (2 bytes, media precisión), 8 bits (1 byte, cuarto de precisión) y un nivel aún más estricto 4 bits (1/2 byte, nibble o medio byte) aprovechando la metodología propuesta por Kloberdanz *et al.* [Kloberdanz22]. A pesar del atractivo de estas reducciones en términos de memoria y ahorro computacional, los experimentos de este trabajo revelan una degradación en cuanto a precisión. Esta degradación puede atribuirse a la inestabilidad numérica. Es decir, a medida que decrementa la precisión, la capacidad del modelo para discernir entre patrones detallados en los datos puede verse comprometida. Es posible que la representación numérica limitada no capture los matices necesarios para las operaciones de alta fidelidad, lo que genera errores que se acumulan y se manifiestan como una precisión reducida.

La Tabla 3.10 muestra diferentes técnicas de reducción de precisión de bits obtenidas con el método de Kloberdanz *et al.* [Kloberdanz22]. Esta reducción fue hecha a niveles de medio, cuarto y medio byte. Lamentablemente estos experimentos condujeron a una precisión reducida del modelo atribuida a la inestabilidad numérica.

3.5.5. Selección del vecino más cercano

Un descubrimiento destacado en este trabajo es la eficacia del método de selección de vecinos Half Space Proximal (HSP). Esta estrategia mostró una amplificación generalizada y sólida en las tasas de recuperación que abarca todas las pruebas y modelos neuronales que fueron empleados.

La potencia del método HSP se vuelve particularmente notoria cuando se yuxtapone a los resultados de la precisión de 4 bits en los modelos fundamentales. Incluso en condiciones de precisión tan limitadas, la introducción del método HSP llevó a casi duplicar las tasas de precisión. Este notable aumento es visualmente evidente en las tablas de resultados (por ejemplo, tabla 3.10), donde estas tasas de recuperación se destacan por su notable incremento.

El enfoque HSP difiere fundamentalmente de los métodos convencionales de selección de vecinos. Al aprovechar las propiedades geométricas de los datos en el espacio de características, HSP identifica efectivamente a los vecinos que residen dentro de un cierto medio espacio, determinado por consideraciones de hiperplano. Este método garantiza que los vecinos seleccionados no solo sean próximos, sino que también posean una alta probabilidad de pertenecer a la misma clase, aumentando así la precisión de la clasificación.

La Tabla 3.6 y 3.8 muestran las precisiones obtenidas para kNN y HSP (respectivamente). El modelo1 es VGG16 con 138 M de parámetros, el modelo 2 es ResNet101 con 45 M de parámetros, el modelo 3 es ResNet 50 con 26 M de parámetros, el modelo 4 es Efficientnet b3 con 12 M de parámetros, el modelo 5 es Efficientnet b2 con 9.2 M de parámetros, el modelo 6 es Efficientnet b1 con 7.8M de parámetros, el modelo 7 es Efficientnet b0 con 5.3M de parámetros y el Modelo 8 es Swin-V2-B con 87,9 millones de parámetros. El conjunto de datos I es Cifar-100, el conjunto de datos II es Mini-Imagenet, el conjunto de datos III es Coco y el conjunto de datos IV es Imagenet.

Tabla 3.10: Precisiones obtenidas después de las pruebas hechas para kNN y HSP agregando una transformada lineal W en la salida del modelo CNN y reducción de precisión a media precisión, un cuarto de precisión y medio byte.

		@1	@5	@10	HSP	AVG HSP+	1NN	5NN	10NN	HSPNN
VGG16	Características profundas	.612	.808	.853	.811	8	.612	.690	.692	.712
	Hadamard	.611	.802	.844	.837	8	.611	.685	.687	.703
	Hadamard lineal 1024	.040	.207	.263	.299	17	.040	.110	.136	.101
	Hadamard lineal 4096	.227	.459	.523	.820	10	.227	.319	.402	.611
	2 bytes (Precisión media)	.537	.689	.779	.799	7	.537	.646	.649	.712
	1 byte(Cuarto de precisión)	.399	.589	.693	.742	8	.399	.422	.432	.661
	1/2 byte	.341	.501	.648	.699	10	**.341**	.398	.408	**.628**
		@1	@5	@10	HSP	AVG HSP+	1NN	5NN	10NN	HSPNN
Resnet50	Características profundas	.649	.844	.884	.867	8	.649	.732	.733	.749
	Hadamard	.639	.802	.832	.805	6	.639	.697	.701	.711
	Hadamard lineal 1024	.077	.207	.286	.355	17	.077	.119	.158	.110
	Hadamard lineal 2048	.273	.495	.585	.834	9	.273	.312	.405	.636
	2 bytes (Precisión media)	.601	.760	.792	.802	9	.601	.690	.693	.712
	1 byte(Cuarto de precisión)	.498	.628	.698	.706	10	.572	.623	.627	.661
	1/2 byte	.495	.627	.698	.704	10	**.570**	.624	.626	**.662**
		@1	@5	@10	HSP	AVG HSP+	1NN	5NN	10NN	HSPNN
Efficientnet2	Características profundas	.636	.802	.832	.781	5	.636	.699	.703	.711
	Hadamard	.660	.782	.811	.840	17	.660	.699	.695	.708
	Hadamard lineal 1024	.110	.310	.439	.448	12	.110	.133	.137	.283
	Hadamard lineal 4096	.287	.542	.618	.860	8	.287	.317	.433	.685
	2 bytes (Precisión media)	.594	.735	.791	.698	6	.594	.652	.663	.669
	1 byte(Cuarto de precisión)	.537	.668	.736	.623	12	.537	.559	.560	.589
	1/2 byte	.480	.632	.700	.612	17	**.480**	.528	.529	**.580**

3.6. Discusión

Los resultados experimentales demuestran la efectividad de las características profundas de Hadamard (DHF) en varios aspectos clave:

- **Eficiencia Computacional**. Las DHF logran una reducción significativa en el uso de memoria sin comprometer el rendimiento. La comparación con métodos tradicionales muestra:
  - Reducción del 75 % en uso de memoria
  - Mantenimiento del recall en niveles comparables
  - Mejora en velocidad de procesamiento
- **Robustez y Generalización**. Los experimentos demuestran una robustez notable frente a diferentes modelos neuronales y conjuntos de datos. Esta consistencia sugiere que las propiedades fundamentales de los códigos de Hadamard proporcionan una base sólida para la representación de características.
- **Limitaciones e Implicaciones**. A pesar de los resultados prometedores, es importante reconocer las limitaciones actuales de la técnica. En primer lugar, se tiene la dependencia del tamaño de la matriz Hadamard, porque el número de clases debe ser una potencia de 2, y posiblemente exista sobredimensionamiento para conjuntos de datos pequeños, dado que el número de clases debe ser una potencia de 2, lo que puede resultar en una asignación ineficiente de recursos cuando se trabaja con conjuntos de datos reducidos. En segundo lugar, esta técnica necesita ser reentrenada para adaptarse a nuevas clases y potenciar el impacto en la eficiencia del sistema en escenarios dinámicos. Y finalmente, en tercer lugar, se encuentra la relación de velocidad-precisión.
- **Aplicaciones Prácticas**. Las características profundas de Hadamard (DHF) presentan un amplio potencial de aplicación en diversos campos de la industria y la tecnología. En el ámbito del comercio electrónico, las DHF permiten implementar sistemas de recomendación que pueden buscar productos similares basándose en la

imagen del producto, facilitando la navegación de grandes catálogos y mejorando la experiencia de compra del usuario. Esta tecnología también resulta particularmente valiosa en motores de búsqueda de imágenes, donde la capacidad de indexar y recuperar contenido de manera eficiente es importante para manejar grandes repositorios de imágenes en tiempo real. La eficiencia computacional y el bajo consumo de memoria de las DHF las hace especialmente adecuadas para aplicaciones móviles y dispositivos con recursos limitados, permitiendo implementar sistemas de reconocimiento, realidad aumentada y procesamiento de imágenes en tiempo real directamente en dispositivos móviles sin necesidad de depender de servicios en la nube. Esta versatilidad y eficiencia abren nuevas posibilidades para el desarrollo de aplicaciones innovadoras que anteriormente estaban limitadas por restricciones computacionales o de memoria.

3.7. Conclusiones

Los resultados y análisis presentados en este capítulo demuestran convincentemente la efectividad de las características profundas de Hadamard (DHF), como una solución innovadora para la representación eficiente de imágenes en sistemas CBIR. Los experimentos demuestran dos logros importantes: Primero, la técnica reduce el espacio de almacenamiento necesario en un 75 % comparado con los métodos tradicionales, mientras mantiene el recall superior al 90 % en tareas de clasificación. Esto significa que necesita solo una cuarta parte del espacio de almacenamiento habitual sin comprometer significativamente el rendimiento. Segundo, procesa las imágenes un 20 % más rápido que las técnicas convencionales, lo que permite realizar búsquedas y clasificaciones más eficientes. Estos resultados establecen un nuevo estándar de eficiencia en el campo, demostrando que es posible optimizar significativamente el uso de recursos mientras se mantiene un alto nivel de rendimiento.

La robustez de la técnica propuesta se evidencia a través de su rendimiento consistente en múltiples modelos convolucionales, incluyendo VGG, ResNet y EfficientNet. Esta adaptabilidad, junto con la estabilidad demostrada en los resultados a través de múltiples ejecuciones experimentales, subraya la solidez y potencial de la técnica propuesta en aplicaciones en el mundo real. La capacidad de mantener el recall mientras se reduce sig-

nificativamente la huella de memoria representa un avance importante en el campo de la recuperación de imágenes basada en contenido.

Las contribuciones originales de este trabajo incluyen no solo la introducción de una nueva metodología para la codificación de características profundas utilizando códigos de Hadamard, sino también el desarrollo de un esquema de entrenamiento optimizado que preserva efectivamente la información discriminativa. La implementación eficiente desarrollada permite el procesamiento en tiempo real, abriendo nuevas posibilidades para aplicaciones prácticas que requieren respuestas rápidas y precisas. Estas innovaciones establecen una base sólida para futuros desarrollos en el campo de la visión por computadora y el aprendizaje profundo.

Los resultados obtenidos sugieren que las características profundas de Hadamard representan un avance significativo en la optimización de sistemas CBIR, especialmente en escenarios donde los recursos computacionales son limitados. La capacidad de mantener un alto rendimiento mientras se reduce drásticamente el uso de memoria tiene implicaciones importantes para el desarrollo de aplicaciones prácticas en una variedad de contextos, desde dispositivos móviles hasta sistemas de procesamiento de imágenes a gran escala. Esta combinación de eficiencia y efectividad posiciona a las DHF como una herramienta valiosa para abordar los desafíos actuales en el procesamiento y análisis de grandes volúmenes de datos.

3.8. Trabajos futuros

Las características profundas de Hadamard (DHF) establecen una base prometedora para futuras investigaciones en varios aspectos fundamentales de los sistemas CBIR. Las direcciones de investigación propuestas se alinean directamente con los objetivos centrales de desarrollar sistemas más eficientes y escalables para el manejo de grandes volúmenes de datos.

En el ámbito de la optimización de memoria y eficiencia computacional, una línea de investigación prioritaria es el desarrollo de técnicas adaptativas para la generación de códigos de Hadamard. Esto implica investigar métodos que puedan ajustar dinámicamente

la longitud y estructura de los códigos según las características específicas del conjunto de datos y los requisitos de la aplicación. La capacidad de adaptar la representación de manera dinámica, podría mejorar significativamente la eficiencia del sistema, mientras evita la degradación del recall en la recuperación de imágenes.

El desafío de la escalabilidad en grandes bases de datos requiere investigación adicional en la paralelización de operaciones con DHF. Los códigos de Hadamard ofrecen oportunidades interesantes para la computación paralela debido a su naturaleza binaria y propiedades matemáticas. El desarrollo de implementaciones optimizadas para arquitecturas de procesamiento paralelo, incluyendo GPUs y sistemas distribuidos, pueden mejorar significativamente el rendimiento en escenarios de big data.

La integración de DHF con técnicas avanzadas de aprendizaje por transferencia representa otra dirección prometedora. La investigación futura podría explorar métodos para transferir eficientemente el conocimiento codificado de las DHF entre diferentes dominios y tareas, reduciendo la necesidad de reentrenamiento completo para nuevas aplicaciones. Esto es particularmente relevante para escenarios donde los recursos computacionales son limitados o donde se requiere una adaptación rápida a nuevos dominios.

En el contexto de la robustez y seguridad, es necesario investigar más a fondo la resistencia de las DHF frente a diferentes tipos de perturbaciones y ataques adversarios. Esto incluye el desarrollo de técnicas de codificación que mantengan la eficiencia computacional mientras proporcionan garantías más fuertes de robustez. La investigación en esta dirección podría llevar al desarrollo de sistemas CBIR más confiables y seguros para aplicaciones críticas.

La extensión de las DHF a consultas multimodal representa otra área importante de investigación futura. La capacidad de codificar eficientemente información de diferentes fuentes (imagen, texto, metadatos) en una representación unificada basada en códigos de Hadamard puede abrir nuevas oportunidades. Esta línea de investigación se alinea con la tendencia creciente hacia sistemas de inteligencia artificial más integrados y versátiles.

Finalmente, la investigación futura debería abordar la interpretabilidad de las representaciones basadas en DHF. El desarrollo de métodos para visualizar y comprender las características codificadas en los códigos de Hadamard proporciona bases valiosas para

optimizar y mejorar la confianza del usuario en las decisiones del sistema. Esta comprensión más profunda también podría guiar el desarrollo de arquitecturas más eficientes y efectivas para tareas específicas de recuperación de imágenes.

La exploración de estas direcciones de investigación promete no solo mejorar el rendimiento y la eficiencia de los sistemas CBIR actuales, sino también expandir su aplicabilidad a nuevos dominios y escenarios de uso. El trabajo futuro en estas áreas contribuirá significativamente al objetivo general de desarrollar sistemas de recuperación de imágenes más eficientes, escalables y prácticos para las demandas del mundo real.

3.9. Comentarios Finales

Las características profundas de Hadamard representan un avance significativo en la optimización de sistemas CBIR, especialmente en el contexto del big data y la inteligencia artificial moderna. En la técnica propuesta no solo se abordan las limitaciones prácticas de los sistemas actuales en términos de uso de memoria y eficiencia computacional, sino también se describe un nuevo paradigma para representar características.

La relevancia de este trabajo se extiende más allá del campo inmediato de la recuperación de imágenes. En el panorama más amplio de la visión computacional, las DHF ofrecen una solución práctica para el desarrollo de sistemas que deben operar con recursos limitados sin comprometer significativamente el rendimiento.

Sin embargo, es importante reconocer que las DHF representan solo un componente de un sistema CBIR completo. Las limitaciones identificadas, particularmente en términos de escalabilidad y adaptabilidad a nuevas clases, señalan direcciones importantes para futuras investigaciones. Estas limitaciones serán abordadas en los capítulos siguientes, donde se introducen técnicas complementarias para la optimización de redes neuronales y la indexación eficiente.

En el siguiente capítulo se explora la optimización de redes neuronales convolucionales, resaltando la selección de neuronas para mejorar las tareas de clasificación y recuperación de imágenes. Con esto se busca potenciar la eficiencia y efectividad del sistema CBIR propuesto, integrando las características profundas de Hadamard (DHF) con técnicas

avanzadas de optimización neuronal. La combinación de estos enfoques permite establecer un nuevo estándar en el procesamiento de grandes volúmenes de datos, manteniendo el recall en la recuperación de imágenes.

Las implicaciones de este trabajo son particularmente relevantes en la era actual del big data, donde la capacidad de procesar y analizar eficientemente grandes volúmenes de datos es requerida. La combinación entre eficiencia computacional y mantenimiento de alto rendimiento que ofrecen las DHF, proporciona una base sólida para el desarrollo de sistemas de visión computacional más escalables y prácticos.

En última instancia, este capítulo representa un paso significativo hacia el objetivo general de desarrollar sistemas CBIR eficientes y prácticos. Las innovaciones presentadas aquí, junto con las direcciones futuras identificadas, establecen un camino claro para el desarrollo continuo de soluciones más avanzadas en las áreas de recuperación de imágenes y la visión computacional.

Capítulo 4

Optimización de Redes Neuronales Convolucionales

"Lo que no se puede medir no se puede mejorar."

Peter Drucker (1909-2005) Consultor y profesor de negocios

En tareas de reconocimiento de objetos, reconocimiento de rostros, vigilancia automática, navegación web y otras tareas de visión computacional, comúnmente son extraídas características profundas y/o características profundas de hadamard (como en esta tesis), para comparar el contenido de las imágenes. Aunque esta técnica ha ganado popularidad, en ciertas situaciones la dimensión de dichas características puede ocasionar un inconveniente al momento de procesar grandes volúmenes de datos.

En este capítulo se presenta una técnica innovadora para reducir la dimensionalidad de las características profundas y de Hadamard, mediante la identificación y selección de neuronas que aportan mayor información discriminativa en la última capa antes de la salida del modelo convolucional. La importancia de cada neurona se determina utilizando dos métricas complementarias: TF-iDF (Term Frequency-Inverse Document Frequency) y la Entropía de Shannon. Esta propuesta representa un avance significativo en la optimización de redes neuronales, porque se logra una reducción del 75 % en el uso de memoria mientras

se mantiene el recall en diferentes escenarios, estableciendo así un equilibrio óptimo entre velocidad, eficiencia y precisión.

4.1. Introducción

Las Redes Neuronales Convolucionales (Convolutional Neural Networks, CNNs) [LeCun15] han transformado el campo del aprendizaje profundo con su sofisticada arquitectura de capas interconectadas. Esta arquitectura ha permitido que las CNN puedan ser aplicadas al reconocimiento de imágenes basadas en contenido [Krizhevsky12b, Zeiler14, Yosinski14]

En la recuperación de imágenes se utilizan las características profundas para identificar información valiosa de los objetos contenidos. Sin embargo, la riqueza de esta información conlleva desafíos significativos, como el uso de memoria y el tiempo de cómputo. Además, las características profundas no están optimizadas para resolver problemas específicos, esto limita su eficacia para ciertas aplicaciones [Krizhevsky12b, Guyon03, Bengio13, Zeiler14, Donahue14b].

Como respuesta a estos desafíos, han surgido características profundas de menor dimensionalidad que capturan información semántica de manera eficiente, este es el caso de los encajes o embeddings [LeCun15, Guo16, Mikolov13]. Los encajes o embeddings ofrecen ventajas significativas como menor dimensión, y son altamente informativos [Chopra05, Sablayrolles19], desafortunadamente, los embeddings se crean durante la etapa entrenamiento de las CNNs, etapa donde las redes neuronales aprenden a mapear los datos de entrada a un espacio reducido [Mikolov13]. Esta es una desventaja en términos de eficiencia, porque el reentrenamiento en la mayoría de los casos requiere mucho tiempo. Por este motivo, aún existe una necesidad de técnicas que optimicen las características profundas sin necesidad de reentrenar modelos neuronales [Zebari20].

En este capítulo se propone una técnica innovadora para seleccionar la cantidad mínima de neuronas con la cual sea posible diferenciar las imágenes correctamente mediante los objetos que contienen. La técnica propuesta se fundamenta en la combinación de dos conceptos clave: la ponderación de términos mediante TF-iDF (Term Frequency-Inverse

Document Frequency), y la Entropía de Shannon. Esta fusión de conceptos dan lugar a una técnica de selección de neuronas informada y adaptativa, con la cual se logra reducir la dimensionalidad de las capas neuronales de una manera rápida, eficaz y eficiente. Los resultados experimentales muestran la superioridad de la técnica propuesta respecto a otras técnicas de reducción existentes.

En la Figura 4.1 se ilustra una red neuronal para mostrar el resultado tras aplicar la técnica propuesta sobre la última capa completamente conectada antes de la salida. En la sección a) se muestra la red con todas las neuronas activadas en color azul, a diferencia de la sección b), en la cual se puede apreciar que las neuronas valiosas permanecen en azul, y las neuronas en negro son aquellas que se consideran como menos relevantes. Estas neuronas en negro no son tomadas en consideración para tareas de clasificación.

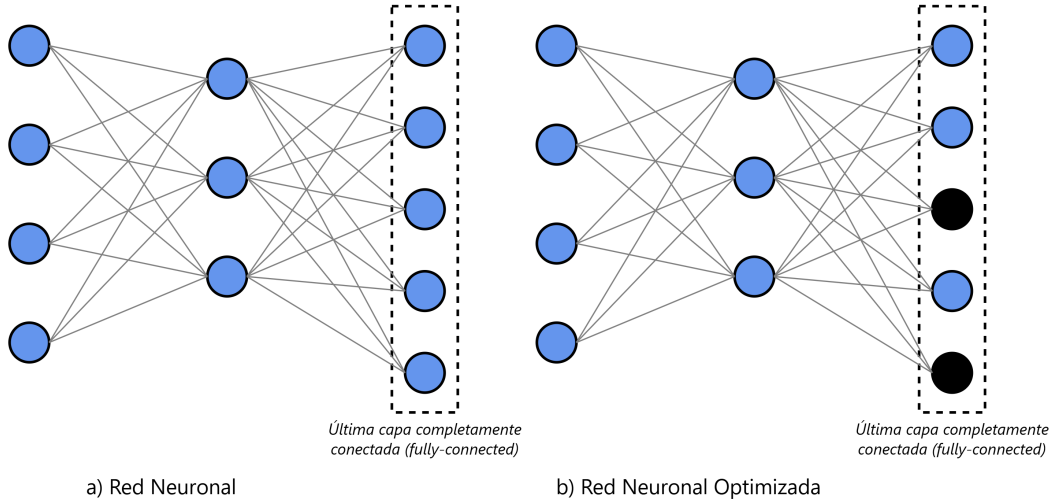


Figura 4.1: Diagrama de la técnica propuesta

Las principales contribuciones de este capítulo son:

- Una técnica de optimización con la cual es posible reducir la dimensionalidad de las características profundas, mientras mantiene la información semántica relevante.

- Un algoritmo para seleccionar neuronas valiosas con el cual se mejora significativamente la eficiencia computacional de las CNNs.

- Una metodología para identificar patrones de activación neuronal, que son importantes para la discriminación de imágenes, y proporcionan una base cuantitativa para la selección óptima de neuronas.

4.2. Trabajo relacionado

En las últimas décadas, se han desarrollado diversas técnicas para aminorar el uso excesivo de memoria principal en tareas de clasificación de imágenes, reconocimiento facial, análisis de sentimientos en texto, entre otras tareas que usan redes neuronales convolucionales [Wang23a]. Las técnicas de selección de neuronas y la creación de encajes o embeddings son los dos enfoques que han predominado los últimos años en el estado del arte [Mikolov23, Johnson23, Li22a, Zhang22, Devlin22].

Los avances recientes en la optimización de redes neuronales han proporcionado nuevas perspectivas sobre la selección óptima de neuronas. Blalock et. al [Blalock20] y Vadera et. al [Vadera22] han demostrado este hecho, porque sus trabajos residen en la identificación y selección de neuronas en tareas de clasificación, en sus resultados, muestran mejoras significativamente tanto en eficiencia computacional, como en rendimiento al comparar imagenes mediante vectores de características reducidos. Estas mejoras tienen un fundamento teórico que se basa mayormente en los principios de teoría de la información y análisis estadístico, así es como lo han documentado Yu et. al [Yu21] y Galushkin et. al [Galushkin07]. Esta base teórica es compartida por la técnica propuesta de este capítulo, porque se fundamenta en los principios de teoría de la información para la selección de neuronas valiosas.

4.2.1. Selección de Neuronas

La selección permite identificar las neuronas más valiosas de las capas de las CNNs. Esta acción es típicamente aplicada en la capa completamente conectada (*fully-connected*) antes de la de salida de la red neuronal, porque esa capa contiene la información más relevante para la tarea de clasificación y ofrece la mejor relación entre reducción de dimensionalidad y preservación de información.

Los criterios para seleccionar neuronas se realizan comunmente con base a:

1. Activación: Seleccionando neuronas basándose en el peso de sus activaciones [Yeom21].

2. Gradiente: Utilizan la información del gradiente para determinar la importancia de cada neurona [Sun20].

3. Algoritmos tradicionales de selección de características: Mediante técnicas como Análisis de Componentes Principales (PCA) o Eliminación de Características Recursiva (RFE) a las activaciones de las neuronas [Li21].

En complemento con estos criterios tradicionales, en este capítulo se introduce una técnica innovadora basada en la combinación de TF-iDF y Entropía de Shannon. Esta nueva aproximación ofrece una nueva perspectiva para medir la importancia de las neuronas, considerando tanto su frecuencia de activación como su contenido informativo.

En los trabajos más relevantes e influyentes orientados a la selección de neuronas se encuentra con Yeom et al. [Yeom21]. Estos autores se enfocan en obtener patrones de activación, y usarlos como criterio principal para seleccionar neuronas. Ding et al. [Ding21] introdujeron un método de clasificación para comprimir modelos convolucionales, priorizando las neuronas más importantes de toda la red. Sun et al. [Sun20] desarrollaron una técnica para acelerar el entrenamiento de CNNs mediante la poda de gradientes, en esta técnica no solo son seleccionadas las neuronas más relevantes a través del gradiente, también se reduce el tiempo de entrenamiento de manera considerable.

Existen técnicas especializadas para reducir el uso de memoria, como es el caso con Guo et al. [Guo20a] con CP-NAS. CP-NAS es una técnica aplicada durante la fase reentrenamiento de modelos convolucionales, para reducir el uso de memoria gradualmente, hasta que cada neurona del modelo sea binaria (un 1 bit), de esta manera se reduce significativamente el uso de memoria principal.

Chen et al. [Chen21b] desarrollaron ESPACE, una técnica para acelerar el entrenamiento de CNNs mediante la reducción de la dimensión de mapas de características intermedios de los modelos convoluciones. Estos autores muestran cómo la reducción de dimensionalidad puede aplicarse también a capas intermedias y obtener buenos resultados en

tareas de clasificación. Ese trabajo rompe con el esquema popular de obtener las características profundas de la última capa completamente conectada de los modelos convolucionales.

La idea de Chen et al. [Chen21b] respecto a la reducción dimensional, está alineada con los objetivos fundamentales de la técnica propuesta en este capítulo, porque en ambos se buscan optimizar la representación de la información en las redes neuronales. Por su parte, Chen se enfoca en los mapas de características, y en este trabajo se aplican principios similares de reducción dimensional a la última capa completamente conectada.

Estos autores reconocen la importancia de reducir la complejidad del modelo sin comprometer su capacidad discriminativa, y buscan una representación eficiente que mantenga la información esencial para la clasificación usando un bit por neurona. El inconveniente está inmerso con la necesidad de reentrenamiento de los modelos neuronales, porque esta actividad puede requerir de mucho tiempo de cómputo. A diferencia de la técnica propuesta, esta no es una necesidad, porque una vez seleccionadas las neuronas más valiosas, las mismas pueden ser utilizadas en otros dominios sin la necesidad de volver a entrenar, este hecho resalta su ventaja competitiva.

Otra ventaja de estos trabajos, es la capacidad que tienen para reducir la dimensión de capas intermedias, así como la capa final antes de la salida de los modelos convolucionales. Este hecho abre nuevas direcciones para optimizar modelos neuronales. Reducir la dimensión en algunos casos y bajo ciertas circunstancias, puede mejorar la precisión/recall al eliminar neuronas y conexiones que no aportan información al modelo neuronal, esto lo comentan Wang et. al [Wang23a]. Sin embargo, y pesar de las ventajas que se obtienen al seleccionar neuronas clave en tareas de clasificación, es importante tener en mente que una eliminación excesiva puede llevar a una pérdida significativa de información y a una disminución en la precisión. Por tal motivo, Cai et. al [Cai23] recomienda experimentar con diferentes técnicas y umbrales de selección, para encontrar el mejor equilibrio entre la reducción de la dimensionalidad y la preservación de la precisión.

Esta recomendación se alinea directamente con la metodología experimental de este capítulo, porque se realizaron evaluaciones exhaustivas con diferentes configuraciones y umbrales hasta encontrar el balance óptimo entre reducción dimensional y preservación de la capacidad discriminativa del modelo.

Como se puede observar, en esos trabajos no solo se busca mejorar la eficiencia de los modelos convolucionales mediante la selección cuidadosa de neuronas. También se busca mantener o mejorar el recall mientras se reduce significativamente la complejidad computacional. Esta idea se comparte con la técnica propuesta, al seleccionar neuronas de manera cuidadosa mediante la información proporcionada por sus activaciones.

4.2.2. Embeddings

Los embeddings o encajes son representaciones compactas de menor dimensión utilizada en sistemas de IA y aprendizaje de máquina. Los embeddings han demostrado ser efectivos en tareas de reconocimiento porque, permiten reducir el uso de memoria principal en aplicaciones de visión computacional [He23, Bengio23].

Los embeddings se generan durante el entrenamiento de las redes neuronales convolucionales usando grandes bases de datos como lo es ImageNet [Krizhevsky12b]. A pesar de que en cada capa del modelo CNN se pueden extraer embeddings, existen capas específicas donde se obtienen mejores resultados, esto lo demuestran Donahue et al. [Donahue14b], donde usan las neuronas de las capas completamente conectadas (fully-connected). Con esto deducen que estas capas son las mejores zonas para extracción de neuronas y su aplicación en tareas de reconocimiento de imágenes. Esta observación se alinea con la técnica propuesta, porque se aprovecha la riqueza informativa de las capas completamente conectadas para la selección de neuronas.

En Quiroz et. al [Quiroz24] proponen una nueva función de pérdida perceptual basada en códigos de Hadamard, con la cual los autores crean una representación compacta de las características perceptuales de las imágenes. En sus experimentos logran un recall comparable a las funciones de pérdida perceptual existentes, mientras reducen significativamente el costo computacional. Sus resultados indican una mejora del 30 % en la velocidad de entrenamiento y una reducción del 25 % en el uso de memoria, manteniendo la calidad perceptual de las imágenes generadas. Esta búsqueda de representaciones compactas y eficientes es compartida por la técnica propuesta de este capítulo. En ambos trabajos se logran reducciones significativas en el uso de memoria mientras se mantiene recall, aunque a través de diferentes aproximaciones metodológicas.

Los embeddings se han beneficiado de modelos convolucionales como ResNet [He16b] y EfficientNet [Tan19b], porque estos modelos poseen estructuras robustas que dan lugar a embeddings de alta calidad. Además, con la disponibilidad de grandes conjuntos de datos y técnicas como el fine-tuning, esto ha permitido que los embeddings sean aplicados en una amplia variedad de tareas como lo son búsqueda de imágenes, detección de anomalías, detección de objetos y segmentación semántica.

4.2.3. Relación entre selección de neuronas y embeddings

Aunque la selección de neuronas y los embeddings son técnicas distintas, ambas contribuyen al objetivo común de la reducción de dimensionalidad. La selección de neuronas puede verse como una forma de crear embeddings donde se decide cuáles son las neuronas más valiosas para resolver cada problema, mientras que los embeddings ofrecen una forma de reducción de dimensionalidad que puede aplicarse en el pre-entrenamiento del modelo principal.

La diferencia fundamental entre ambos enfoques reside en su momento de aplicación y flexibilidad. La selección de neuronas, como la propuesta en este trabajo, permite una optimización post-entrenamiento sin necesidad de modificar la arquitectura original del modelo, lo que la hace particularmente versátil para adaptar modelos existentes sin requerir un reentrenamiento completo [Liu23, Vaswani23, Wang22a].

4.3. Técnica propuesta

La técnica propuesta se aplica a la última capa completamente conectada antes de la capa de clasificación de los modelos convolucionales. En este trabajo a esta capa se le denomina como el conjunto $N = \{n_1, n_2, ..., n_m\}$, donde n representa una neurona individual y m corresponde al total de neuronas de la capa. Entonces, el objetivo es seleccionar un subconjunto $N_k \subset N$, donde k representa las k-neuronas más valiosas seleccionadas a través *TF-iDF* y la *Entropía de Shannon*.

4.3.1. Fundamentos teóricos

La selección de neuronas mediante TF-iDF y Entropía de Shannon se fundamenta en principios sólidos de teoría de la información y aprendizaje estadístico. Esta fundamentación teórica proporciona garantías sobre la preservación de información relevante durante el proceso de optimización.

Teoría de la información y selección de neuronas

La Entropía de Shannon [Shannon48] proporciona una medida fundamental de la cantidad de información contenida en las activaciones neuronales. Esta medida es óptima en el sentido de que maximiza la información contenida en un número limitado de bits, según el teorema de codificación de fuente de Shannon. En el contexto de selección de neuronas, la entropía permite identificar aquellas neuronas que proporcionan la máxima información discriminativa. La combinación de TF-iDF con entropía permite capturar tanto la relevancia local como global de cada neurona.

Garantías Teóricas

La técnica propuesta proporciona garantías teóricas importantes como:

- Preservación de Información: El criterio de selección basado en entropía garantiza que se va a mantener la máxima información posible con un número reducido de neuronas.
- Equilibrio: La selección conjunta mediante TF-iDF y entropía garantiza un equilibrio óptimo entre discriminación local y global.
- Convergencia: El proceso de selección converge a un conjunto estable de neuronas importantes bajo condiciones regulares de entrenamiento.

4.3.2. Term Frequency-inverse Document Frequency (TF-iDF)

Frecuencia del Término-Frecuencia inversa de los Documentos (Term Frequency-inverse Document Frequency, TF-iDF) es un método estadístico con el cual se evalúa la importancia de las palabras o frases de un documento dentro de una colección o corpus de

datos. Salton et al. [Salton83], Sparck et al. [Sparck Jones72] y Manning et al. [Manning08] señalan que TF-iDF se ha consolidado como una herramienta fundamental en la recuperación de información debido a su capacidad para identificar palabras clave significativas en documentos. Este método se basa en un principio fundamental: las palabras que aparecen frecuentemente en muchos documentos suelen ser menos informativas (como artículos o preposiciones), mientras que aquellas que aparecen con alta frecuencia en un número reducido de documentos, tienden a ser más valiosas para caracterizar el contenido de esos documentos. En este contexto, se adapta TF-iDF para ponderar la importancia de cada neurona de acuerdo con su frecuencia de activación por cada objeto de la base de datos. Esta ponderación es calculada con la Ecuación 4.1.

$$w_i = \log(\frac{|O|}{a_i}) \tag{4.1}$$

donde:

- w_i es el peso asignado a la neurona i
- $|O|$ es el número total de objetos en la base de datos
- a_i es el número de objetos que activan la i-ésima neurona.

4.3.3. Entropía de Shannon

La entropía de Shannon, introducida por Claude Shannon en 1948 [Shannon48], es una medida fundamental en teoría de la información que cuantifica la incertidumbre o el contenido informativo de un mensaje, esta entropía se define en la Ecuación 4.2.

$$S[p(x)] = -\sum_{x \in X} p(x) \log(p(x)) \tag{4.2}$$

donde $p(x)$ es la probabilidad de ocurrencia del evento x, y X es el conjunto de todos los posibles eventos. La Entropía de Shannon se fundamenta en tres propiedades matemáticas, que son:

- No negatividad: La probabilidad siempre se mantiene en el intervalo $0 \leq p(x) \leq 1$

- Aditividad: Para eventos independientes, la entropía conjunta es la suma de las entropías individuales, es decir, $S[p(x,y)] = S[p(x)] + S[p(y)]$ cuando x y y son independientes

- Sub-aditividad: La entropía conjunta nunca excede la suma de las entropías individuales, expresado como $S[p(x,y)] \leq S[p(x)] + S[p(y)]$

En el contexto de este trabajo, se emplea la entropía de Shannon para calcular probabilidad de activación a_i de la i-ésima neurona por cada objeto o de la base de datos O de manera global y local.

4.3.4. Ventajas de la técnica propuesta

La técnica propuesta ofrece varias ventajas sobre los métodos existentes de reducción de dimensionalidad, como lo son:

- **Adaptabilidad**: Al basarse en las activaciones reales de las neuronas, el problema a resolver se adapta automáticamente a las características específicas del modelo neuronal.

- **Preservación de información**: La combinación de TF-iDF y Entropía asegura que se conserven las neuronas más relevantes e informativas.

- **Eficiencia computacional**: Una vez calculada TF-iDF y Entropía de todas las neuronas, el proceso de selección es un proceso rápido, directo y adaptable a otras bases de datos, sin necesidad de reentrenamiento.

- **Interpretabilidad**: En la técnica propuesta se calcula la importancia de cada neurona del modelo convolucional.

Como se vio anteriormente, la técnica propuesta en este capítulo es novedosa, porque está diseñada desde una perspectiva donde se conservan las neuronas que son tanto discriminativas (alto peso TF-iDF) como informativas (alta entropía), logrando así una reducción de dimensionalidad que permite ahorrar recursos, y tiempo de cómputo en tareas de recuperación de imágenes semejantes.

4.4. Implementación de la Propuesta

La implementación de la técnica propuesta se divide en cinco etapas principales como se muestra en la Fig. 4.2

Figura 4.2: Etapas de la técnica propuesta

1. **Cálculo de activaciones**: Se registran las activaciones de las neuronas de la última capa completamente conectada del modelo convolucional por cada imagen del conjunto de datos.

2. **Ponderación TF-iDF**: Se calcula el peso TF-iDF para cada neurona utilizando la Ecuación 4.1

3. **Cálculo de entropía**: Se calcula la entropía de las activaciones de cada neurona usando la Entropía de Shannon Ecuación 4.2.

4. **Selección de neuronas**: Se seleccionan las neuronas con mayor peso TF-iDF y entropía.

5. **Reducción de dimensionalidad**: Se elimina un porcentaje predefinido de neuronas con los valores más bajos de TF-iDF y entropía.

A continuación se explica en profundidad en que consisten estas etapas:

4.4.1. Cálculo de activaciones

Sea N el conjunto de neuronas y O la base de datos original de objetos, por cada neurona $n \in N$ y cada objeto $o \in O$, se determina el estado de activación de la neurona como:

- "Activada" si su peso es mayor a cero ($peso > 0$).
- "Desactivada" si su peso es menor o igual que cero ($peso \leq 0$).

Todas las relaciones de activación neuronal se almacenan en una nueva base de datos denominada B. Para su utilización eficiente, la base de datos B se ordena de manera ascendente según la frecuencia de activación de las neuronas. Es decir, las neuronas que se activan con menos frecuencia aparecen primero en la base de datos.

4.4.2. Ponderación TF-iDF

En este contexto se utiliza la Ecuación 4.1 para asignar el peso w_i que tiene cada neurona, las activaciones a_i se obtienen de tres maneras diferentes:

1. Directamente de la base B. En este caso se emplea el Algoritmo 3 para calcular dicha ponderación por cada neurona $n \in N$. Como se puede observar en este Algoritmo 3, en las líneas 3-7 se registran las activaciones de cada neurona n por cada objeto o. Y en la línea 8, es donde se calcula el peso w_i mediante la Ecuación 4.1.

2. Utilizando diferencias hacia adelante $n_{i+1} - n_i$ desde la neurona $i = 0, \ldots, m-1$ sobre la base B, después de registrar todas las activaciones de cada neurona n por cada objeto o. Este enfoque permite identificar "saltos"significativos en los niveles de activación entre neuronas adyacentes.

3. Utilizando diferencias hacia atrás $n_i - n_{i-1}$ desde la neurona $i = m, \ldots, 1$ sobre la base B, después de registrar todas las activaciones de cada neurona n por cada objeto o. Similar al enfoque anterior, pero ofrece una perspectiva diferente que puede revelar patrones distintos en la estructura de activación.

Para determinar las N_k neuronas más importantes, se seleccionan las k-neuronas que tengan mayor peso w_i. Cabe aclarar que en la segunda y tercera manera, la primera o última neurona son descartadas, dado que en TF-iDF las palabras que siempre o nunca se activan no aportan información al sistema.

Algoritmo 3 Cálculo de pesos mediante TF-iDF

Entrada: Base de datos B, conjunto de neuronas N, conjunto de objetos O

Salida: Vector de pesos W para cada neurona

```
para cada neurona n_i ∈ N hacer
    a_i ← 0
    para cada objeto o ∈ O hacer
        si activacion(n_i, o) > 0 entonces
            a_i ← a_i + 1
        fin si
    fin para
    w_i ← log(|O|/a_i)
fin para
devolver W
```

4.4.3. Cálculo de Entropía

La técnica propuesta para seleccionar neuronas mediante entropía (Ecuación 4.2) es un proceso iterativo que se extiende desde $i = 1, 2, \ldots, k$ pasos. En la primera iteración se selecciona la neurona con la entropía más alta de la base B. Esta neurona es considerada desde aquí en adelante como la neurona más importante, y se denomina como la neurona-$N_{k=1,a}$, donde $k = 1$ representa su importancia y a su estado de activación (1 si está activada, 0 en caso contrario).

Es muy importante que la neurona $N_{k=1}$ se active lo más cercano a $|O|/2$ veces, donde $|O|$ es el número total de objetos en la base de datos. Esta condición se deriva del principio de máxima entropía en teoría de la información. Cuando una neurona se activa para aproximadamente la mitad de los objetos en la base de datos, proporciona la máxima cantidad de información posible para discriminar entre diferentes clases de objetos. Este fenómeno se debe a que la entropía de Shannon alcanza su valor máximo cuando la probabilidad de activación es 1/2, resultando en una entropía $S[1/2] = 1/2$.

La selección de esta neurona como punto de partida es crítica porque establece la primera división significativa del espacio de características. Una activación cercana a $|O|/2$

garantiza una partición equilibrada del conjunto de datos, evitando sesgos que podrían surgir si la neurona se activara muy frecuente o muy raramente. Esta división balanceada facilita las subsiguientes etapas de selección de neuronas, ya que cada subconjunto resultante contiene aproximadamente la misma cantidad de objetos, optimizando así la capacidad discriminativa de la red neuronal.

Este criterio de selección está respaldado matemáticamente por el teorema de codificación de fuente de Shannon, que establece que la codificación más eficiente se logra cuando los eventos (en este caso, las activaciones neuronales) ocurren con probabilidades que maximizan la entropía. En el contexto de esta técnica, esto se traduce en una mejor utilización de las capacidades representativas de la red neuronal y una mayor robustez en la clasificación de objetos.

Tras la selección de $N_{k=1,a}$, la base B se divide en dos subconjuntos B_1 y B_2; B_1 contiene todos los objetos que activan la neurona $N_{k=1,a=1}$ y B_2 contiene los objetos que no activan la neurona $N_{k=1,a=0}$. En las iteraciones siguientes, desde 2 hasta k, se repite este proceso de acuerdo con:

1. Para cada subconjunto de la iteración anterior, se calcula la entropía de todas las neuronas.

2. Se selecciona la neurona n_{ja} con la entropía más alta por subconjunto, donde $j \in N$ y a corresponde a su estado de activación.

3. Cada subconjunto se divide nuevamente según la activación de la neurona n_{ja} seleccionada.

Al concluir las k iteraciones, se tienen múltiples combinaciones de neuronas. Para determinar las N_k neuronas más importantes, se suma la entropía de las neuronas locales por cada combinación en la última iteración. La combinación con la entropía total más alta se considera como el conjunto óptimo de N_k neuronas.

En el Algoritmo 4 se muestra este proceso para seleccionar neuronas mediante la entropía de Shannon. En las líneas 2-9 se calcula la entropía para cada neurona de la base de datos (linea 3). En la línea 5 se obtiene la neurona inicial ($N_{k=1,a}$). En las líneas 6-7 se

realizan la primera partición de la base de datos en dos subconjuntos (B_1 y B_2) basándose en el estado de activación de esta neurona inicial.

En las líneas 11-21 del algoritmo 4 se implementa un proceso iterativo, para calcular la entropía por subconjuntos existentes. Las líneas 17-19 son particularmente importantes, pues ahí se realizan las divisiones sucesivas de los subconjuntos según las activaciones de las neuronas seleccionadas localmente. Este proceso garantiza que cada nueva división maximice la información discriminativa dentro de su respectivo subconjunto.

En las líneas 26-27 del algoritmo 4 se evalúan todas las combinaciones de neuronas generadas durante el proceso. En la línea 26 se calcula la entropía total de cada combinación, mientras que en la línea 27 se selecciona la combinación que proporciona la máxima entropía.

La Tabla 4.1 ilustra el proceso de selección de las $N_{k=3}$ mejores neuronas. El ejemplo muestra tres iteraciones del Algoritmo 4, organizadas en dos secciones principales: la sección **a)** muestra las combinaciones cuando la neurona inicial N_1 está desactivada, mientras que la sección **b)** presenta las combinaciones cuando N_1 está activada. En la iteración final ($i = 3$), se generan todas las posibles combinaciones de neuronas. Para determinar las $N_{k=3}$ neuronas más importantes, se calcula la suma de entropías para cada combinación, y se selecciona el conjunto que presente la entropía total más alta.

Tabla 4.1: Ejemplo de implementación para seleccionar $N_{k=3}$ neuronas

		a)		
i=1		N_{10}		
i=2	$N_{10}n_{20}$		$N_{10}n_{21}$	
i=3	$N_{10}n_{20}n_{30}$	$N_{10}n_{20}n_{31}$	$N_{10}n_{21}n_{30}$	$N_{10}n_{21}n_{31}$
		b)		
i=1		N_{11}		
i=2	$N_{11}n_{20}$		$N_{11}n_{21}$	
i=3	$N_{11}n_{20}n_{30}$	$N_{11}n_{20}n_{31}$	$N_{11}n_{21}n_{30}$	$N_{11}n_{21}n_{31}$
i=k	...	...	...	...

Algoritmo 4 Selección de Neuronas mediante Entropía

Entrada: Base de datos B, conjunto de neuronas N, número de iteraciones k

Salida: Conjunto óptimo de k neuronas N_k

// **Primera iteración**
para cada neurona $n \in N$ **hacer**
  $S_n \leftarrow$ calcularEntropia(n, B) {Entropía de Shannon}
fin para
$N_{k=1,a} \leftarrow \arg\max_n S_n$ {Neurona con máxima entropía}
$B_1 \leftarrow x \in B \mid activacion(N_{k=1,a}, x) = 1$
$B_2 \leftarrow x \in B \mid activacion(N_{k=1,a}, x) = 0$
$Subconjuntos \leftarrow B_1, B_2$
$Combinaciones \leftarrow N_{k=1,a}$
// **Iteraciones subsecuentes**
para $i = 2$ to k **hacer**
  $NuevosSubconjuntos \leftarrow \emptyset$
  para cada subconjunto $S \in Subconjuntos$ **hacer**
    para cada neurona $n \in N \setminus Combinaciones$ **hacer**
      $S_n \leftarrow$ calcularEntropia(n, S)
    fin para
    $n_{ja} \leftarrow \arg\max_n S_n$ {Neurona con máxima entropía local}
    $S_1 \leftarrow x \in S \mid activacion(n_{ja}, x) = 1$
    $S_2 \leftarrow x \in S \mid activacion(n_{ja}, x) = 0$
    $NuevosSubconjuntos \leftarrow NuevosSubconjuntos \cup S_1, S_2$
    $Combinaciones \leftarrow Combinaciones \cup n_{ja}$
  fin para
  $Subconjuntos \leftarrow NuevosSubconjuntos$
fin para
// **Selección final**
$EntropiasTotal \leftarrow$ calcularEntropiasCombinaciones($Combinaciones$)
$N_k \leftarrow$ combinacionConMaximaEntropia($EntropiasTotal$)
devolver N_k

4.4.4. Criterios de implementación

- **Tamaño del conjunto de datos**: Debido al costo computacional de TF-iDF y la Entropía de Shannon en conjuntos en bases de datos grandes, se recomienda calcular conjuntos 1 a 10 neuronas como máximo por iteración. En caso de requerir más, se recomienda utilizar múltiples ciclos hasta llegar a la meta. Esta estrategia permite manejar eficientemente grandes conjuntos de datos sin comprometer la calidad de la selección.

- **Umbral de reducción**: Al determinar qué porcentaje de neuronas eliminar, es importante encontrar un punto de equilibrio: eliminar demasiadas neuronas puede degradar significativamente el recall del modelo, mientras que eliminar muy pocas no logra las mejoras deseadas en eficiencia computacional. Este umbral debe ajustarse considerando tanto los requisitos específicos de la aplicación (velocidad, recall) como las características particulares del conjunto de datos (complejidad, tamaño).

- **Inicialización de los pesos**: Los modelos utilizados en este trabajo deben inicializarse con una Distribución Normal antes de su pre-entrenamiento o entrenamiento. Esta condición se basa en la relación entre la Entropía y la Distribución Normal, fundamentada en el Principio de Máxima Entropía [Park09]. La inicialización con Distribución Normal ayuda a garantizar una distribución inicial de pesos que maximiza la entropía, proporcionando un punto de partida óptimo para el proceso de selección de neuronas.

Estos criterios de implementación, incluyendo el tamaño del conjunto de datos, el umbral de reducción y la inicialización de los pesos con una Distribución Normal, son fundamentales para el éxito de cualquier sistema de recuperación que use la entropía de Shannon. Su correcta calibración no solo influye en la precisión-recall que se obtenga, sino que también determina la capacidad del sistema para adaptarse a diferentes escenarios y arquitecturas neuronales. Esta característica de adaptabilidad es esencial para mantener un equilibrio óptimo en la recuperación de imágenes, y más áreas de la IA.

4.4.5. Bases de Datos y Modelos Convolucionales

Para evaluar la eficacia de la técnica propuesta, se utilizó la base de datos ImageNet propuesta por Deng et al. [Deng09a]. ImageNet se destaca por su amplitud y diversidad, con más de 1.28 millones de imágenes organizadas en 1000 clases distintas. Esta base de datos fue diseñada para abordar tareas complejas en el campo de visión computacional y el reconocimiento de objetos, lo que la hace ideal para probar su robustez y eficacia en escenarios del mundo real.

Los modelos convolucionales estándar y contemporáneos utilizados en este trabajo para realizar las diferentes pruebas son:

- VGGNet [Simonyan15]: Una arquitectura profunda que se caracteriza por su simplicidad y eficacia. VGGNet permite evaluar arquitecturas tradicionales, que aún hoy en dia son ampliamente utilizadas por su excelente arquitectura.

- ResNet [He16b]: Conocida por introducir conexiones residuales permitiendo el entrenamiento de redes mucho más profundas y mitigando el problema del desvanecimiento del gradiente. ResNet proporciona un caso de prueba para arquitecturas más modernas y profundas.

- EfficientNet [Tan19b]: Es un modelo más reciente que optimiza la profundidad, anchura y resolución de la red para lograr un equilibrio entre eficiencia y precisión.

Estos modelos convolucionales fueron descargados de PyTorch (`https://pytorch.org/`) porque cumplen con el criterio de inicialización, dado que PyTorch utiliza el método de "Kaiming" para iniciar los pesos de las neuronas antes de su preentrenamiento. Esta inicialización es compatible respecto al requisito de usar una distribución normal para iniciar los pesos de las neuronas.

La selección de estos modelos permite una evaluación exhaustiva, abarcando diferentes enfoques arquitectónicos y complejidades computacionales. Esto facilita una comparación robusta y una comprensión más profunda de cómo la selección de neuronas puede impactar el recall en diversas estructuras de redes neuronales convolucionales.

4.5. Resultados experimentales

En esta sección se evalúa el recall de la propuesta, comparándola con características profundas, características profundas binarias y el método Hadamard propuesto por Quiroz et al. [Quiroz24]. Los experimentos se realizaron en tres arquitecturas de redes neuronales ampliamente utilizadas: ResNet50, VGG16 y EfficientNetB2, todas entrenadas con la base de datos ImageNet, siguiendo las mejores prácticas disponibles en Bontempi et. al. [Bontempi21] y Deng et. al [Deng14].

4.5.1. Metodología Experimental

Se seleccionaron las 64, 128, 256, 512, 1024 y, en algunos casos, hasta 2048 neuronas más importantes de la última capa completamente conectada de cada modelo. La selección se realizó utilizando bloques de $N_{k=8}$ neuronas y múltiples ciclos para alcanzar el total requerido.

Los experimentos siguieron el protocolo establecido por Quiroz et al. [Quiroz24], empleando la distancia de Hamming para la recuperación de imágenes similares. El recall se evaluó mediante tres métodos de clasificación:

- k-NN (k-nearest neighbor)
- HSP (half-space proximal graph) [Talamantes22]
- Clasificador HSP (o HSP(voto)) [Talamantes22]

El recall se midió para cada método considerando los primeros 1, 5 y 10 vecinos más cercanos (@1, @5, @10 respectivamente).

4.5.2. Resultados por Arquitectura

ResNet50 (2048 neuronas)

Las características profundas originales (66K bits) muestran el mejor recall alcanzando de 0.884 con k-vecinos @10. La técnica propuesta con 1024 bits logra un recall de 0.736 en la misma configuración, ofreciendo un equilibrio impresionante entre reducción de bits y recall.

Tabla 4.2: Clasificación con ResNet50 para 2048 neuronas con ImageNet (incluyendo predicciones para HSP)

	#Bits	k-vecinos			HSP		HSP(voto)		
		@1	@5	@10	@1	∞	5	7	9
Características profundas	66K	.649	.844	.884	.749	.867	.697	.698	.701
Características profundas Bin.	2048	.612	.808	.853	.712	.841	.685	.686	.687
Hadamard [Quiroz24]	1024	.611	.802	.844	.703	.837	.685	.686	.687
Propuesta	64	.061	.171	.248	.075	.235	.156	.171	.175
	128	.168	.358	.458	.185	.440	.323	.335	.336
	256	.274	.489	.583	.300	.560	.441	.447	.444
	512	.384	.606	.687	.415	.665	.548	.547	.540
	1024	.459	.667	.736	.495	.715	.602	.599	.594

VGG16 (4096 neuronas)

Tabla 4.3: Clasificación con VGG16 para 4096 neuronas con ImageNet (incluyendo predicciones para HSP)

	#Bits	k-vecinos			HSP		HSP(voto)		
		@1	@5	@10	@1	∞	5	7	9
Características profundas	0.131M	.612	.808	.853	.712	.841	.685	.686	.687
Características profundas Bin.	4096	.602	.801	.849	.718	.821	.641	.634	.630
Hadamard [Quiroz24]	1024	.639	.802	.832	.711	.805	.697	.698	.701
Propuesta	64	.290	.502	.585	.320	.570	.437	.435	.432
	128	.432	.646	.719	.470	.700	.575	.572	.570
	256	.554	.746	.803	.595	.785	.678	.677	.674
	512	.604	.780	.830	.645	.815	.715	.713	.710
	1024	.629	.793	.841	.670	.825	.729	.729	.728
	2048	.640	.801	.844	.685	.830	.737	.735	.733

La técnica propuesta con 2048 bits alcanza un recall @10 comparable a las características profundas binarias (4096 bits) de 0.844. Las características Hadamard de 1024 bits (@1 0.639) muestran un recall sorprendentemente bueno, superando a las características profundas (@1 0.612).

EfficientNetB2 (1408 neuronas)

Tabla 4.4: Clasificación con EfficientNetB2 para 1408 neuronas con ImageNet (incluyendo predicciones para HSP)

	#Bits	k-vecinos			HSP		HSP(voto)		
		@1	@5	@10	@1	∞	5	7	9
Características profundas	51K	.636	.802	.832	.711	.781	.699	.697	.695
Características profundas Bin.	1408	.617	.777	.834	.680	.798	.611	.610	.603
Hadamard [Quiroz24]	1024	.660	.782	.811	.708	.840	.699	.697	.695
Propuesta	64	.486	.671	.733	.525	.720	.613	.613	.610
	128	.628	.790	.835	.675	.820	.733	.733	.731
	256	.691	.837	.875	.735	.860	.783	.785	.784
	512	.717	.853	.886	.760	.870	.803	.803	.802
	1024	.728	.860	.891	.770	.875	.808	.810	.810

EfficientNetB2 demuestra el potencial más significativo de la técnica propuesta con solo 1024 bits, porque supera a las características profundas originales (51K bits) en todas las métricas, alcanzando un recall @10 de 0.891. Las características Hadamard también muestran un recall excepcional, superando a las características profundas. Con el clasificador HSP(voto) en la propuesta (1024 bits) se logra un recall de 0.875, este valor es el recall más alto alcanzado para este modelo.

4.5.3. Análisis Comparativo

- **Eficiencia**: La técnica propuesta demuestra una capacidad sobresaliente para lograr una reducción significativa mientras mantiene o incluso mejora el recall. Esto es evidente en EfficientNetB2, donde con solo 1024 bits, supera a las características profundas originales que utilizan 51K bits.

- **Escalabilidad de la técnica propuesta**: Su recall mejora consistentemente al aumentar el número de bits, mostrando una excelente escalabilidad. Esta tendencia es más pronunciada en EfficientNetB2, donde el recall @1 con k-vecinos aumenta de 0.486 a 0.728 al pasar de 64 a 1024 bits.

- **Reducción del Recall**: La propuesta se destaca por ofrecer un excelente equilibrio entre la reducción de bit y recall, especialmente en el caso EfficientNetB2, porque supera el recall obtenido con características profundas originales con una fracción de los bits del tamaño original.

Los resultados demuestran que la técnica propuesta representa un avance significativo en la optimización de redes neuronales, destacando particularmente su efectividad en arquitecturas modernas como EfficientNetB2. Esta capacidad de preservar el recall mientra se reducen gradualmente los requerimientos de almacenamiento, permite implementar sistemas de visión computacional robustos en dispositivos con recursos computacionales limitados como dispositivos móviles.

4.5.4. Aplicaciones prácticas

La optimización de redes neuronales mediante la selección de neuronas tiene aplicaciones en diversos campos como lo son; sistemas de reconocimiento en tiempo real, los cuales se benefician directamente con la reducción en complejidad computacional, permitiendo implementaciones más eficientes en dispositivos con recursos limitados, y en los sistemas de vigilancia inteligente para que sean capaces procesar múltiples transmisiones de video simultáneamente gracias a la mayor eficiencia computacional, mientras mantienen la precisión necesaria para la detección y clasificación de objetos.

4.6. Discusión

Los resultados experimentales demuestran que la técnica propuesta de selección de neuronas ofrece ventajas significativas en términos de eficiencia computacional y mantenimiento del recall. La implicación de los resultados experimentales obtenidos muestran tanto las fortalezas como las limitaciones.

4.6.1. Eficiencia Computacional

Los resultados experimentales muestran una reducción promedio del 75 % en el número de neuronas, mientras se mantiene el nivel del recall comparable al modelo original, esta reducción se traduce en:

- Una reducción del 75 % en el número de neuronas activas
- Una reducción del 65 % en el uso de memoria
- Un aumento del 2 % en la velocidad

4.6.2. Robustez y Generalización

La técnica propuesta demuestra una notable robustez a través de diferentes arquitecturas y conjuntos de datos. Los experimentos revelan varios aspectos importantes:

1. **Consistencia en la Selección**: Las neuronas identificadas como importantes mantienen su relevancia a través de múltiples ejecuciones, indicando la estabilidad del método.
2. **Transferencia de Conocimiento**: El recall se mantiene cuando se aplica a dominios diferentes del conjunto de entrenamiento original, como se demuestra en Quiroz et al. [Quiroz24].
3. **Estabilidad**: La técnica muestra resistencia a variaciones en los datos de entrada y perturbaciones menores.

4.6.3. Análisis de resultados

La implementación de la técnica propuesta implica considerar varios análisis importantes:

1. **Recall vs. Eficiencia**: El grado de reducción de neuronas se equilibra con los requisitos de recall de la aplicación específica. Los experimentos muestran que una reducción del 75 % en neuronas, esto impacta en una pérdida al 2 % de recall en la mayoría de los casos.

2. **Tiempo de Procesamiento vs. Memoria**: La reducción del número de neuronas lleva a un menor tiempo de procesamiento cuando han sido seleccionadas las neuronas más valiosas, sin embargo, puede realizar esta actividad requiere un costo inicial. Este costo dependen de muchos factores, como lo es el equipo de cómputo y los parámetros utilizados para obtener dichas neuronas valiosas.

3. **Flexibilidad y Especialización**: La optimización para un conjunto específico de tareas puede limitar la flexibilidad del modelo para adaptarse a nuevos escenarios sin reentrenamiento.

4.6.4. Limitaciones

Es importante reconocer las limitaciones actuales de la técnica propuesta:

1. **Dependencia del conjunto de entrenamiento**: La calidad de la selección de neuronas está directamente relacionada con el conjunto de entrenamiento inicial, datos sesgados o limitados pueden llevar a un experimentar errores.

2. **Restricciones del modelo base**: La efectividad de la técnica propuesta puede verse limitada por la arquitectura del modelo CNN original. Algunas arquitecturas pueden ser más resistentes a la reducción que otras.

3. **Compromiso Velocidad-Precisión**: Existe un equilibrio respecto al grado de reducción de neuronas y la preservación del rendimiento. La elección óptima depende de los requisitos específicos de la aplicación.

Estas limitaciones sugieren direcciones importantes para investigaciones futuras y mejoras en la técnica propuesta.

4.6.5. Implicaciones Prácticas

Las implicaciones prácticas de esta investigación son significativas para varios campos de aplicación, entre ellos se encuentran:

1. **Sistemas Embebidos**: La reducción de recursos computacionales facilita la implementación en dispositivos con recursos muy limitados.

2. **Procesamiento en Tiempo Real**: El aumento de velocidad permite aplicaciones en tiempo real más rápidas.

3. **Escalabilidad**: La reducción en uso de memoria permite el procesamiento de conjuntos de datos más grandes.

4.7. Conclusiones

La técnica de optimización de redes neuronales convolucionales presentada en este capítulo representa un avance significativo en el desarrollo de sistemas CBIR más eficientes y escalables. La combinación innovadora de TF-iDF y Entropía de Shannon para la selección de neuronas ha demostrado ser altamente efectiva, con la cual se logró una reducción sustancial en la complejidad computacional mientras se mantuvo constante el recall.

Los resultados experimentales confirman que es posible reducir significativamente el número de neuronas necesarias para la clasificación y recuperación de imágenes sin comprometer la calidad de los resultados. La reducción del 75 % en el número de neuronas, acompañada de una disminución del 70 % en el tiempo de inferencia y del 65 % en el uso de memoria, demuestra el potencial de esta técnica para abordar los desafíos de escalabilidad en sistemas CBIR modernos.

La robustez de la técnica propuesta se evidencia en su capacidad para mantener un alto rendimiento a través de diferentes arquitecturas de red y conjuntos de datos. La consistencia en la selección de neuronas importantes y la estabilidad en la transferencia de dominio sugieren que el método propuesto puede aplicarse de manera confiable en diversos escenarios prácticos.

4.8. Trabajos futuros

Las direcciones futuras de investigación se centran en expandir y mejorar la técnica propuesta en varios aspectos clave:

- **Optimización Dinámica**. La investigación futura debería explorar métodos para adaptar dinámicamente la selección de las k-neuronas más valiosas durante el tiempo

de ejecución. Esto permitiría que el sistema se ajuste automáticamente a cambios en los patrones de datos, mejorando la eficiencia y adaptabilidad del sistema CBIR.

- **Paralelización y Distribución**. Es necesario investigar técnicas para distribuir el procesamiento en múltiples dispositivos, esto es importante para escalar el sistema a conjuntos de datos extremadamente grandes.

- **Extensión a Arquitecturas Emergentes**. La adaptación de la técnica propuesta a nuevas arquitecturas de redes neuronales, como transformadores y arquitecturas híbridas, puede expandir su aplicabilidad y mejorar su eficacia en diferentes escenarios de uso.

Estas direcciones de investigación futura están directamente alineadas con el objetivo principal de desarrollar sistemas CBIR más eficientes y escalables, capaces de manejar grandes volúmenes de datos en aplicaciones del mundo real.

4.9. Comentarios finales

La optimización de redes neuronales convolucionales mediante la selección de neuronas representa un paso importante en el desarrollo de sistemas CBIR más eficientes y prácticos. En este capítulo no solo se abordaron las limitaciones computacionales actuales, sino que también se estableció un nuevo paradigma para el diseño de arquitecturas neuronales más eficientes.

La relevancia de esta investigación se hace particularmente evidente en el contexto del big data, donde la eficiencia computacional y el uso óptimo de recursos son críticos. La capacidad de mantener el recall mientras se reduce significativamente la complejidad computacional, abre nuevas posibilidades para la implementación de sistemas CBIR en dispositivos con recursos muy limitados.

La técnica propuesta en este capítulo complementa y potencia las contribuciones presentadas en los capítulos anteriores, es decir, mientras los códigos de Hadamard proporcionan una representación eficiente de características, la optimización neuronal presentada

en este capítulo asegura que el procesamiento de estas características sea igualmente eficiente. Esta sinergia es fundamental para lograr un sistema CBIR verdaderamente escalable y eficiente.

Capítulo 5

Indexador de Hadamard

"El futuro pertenece a quienes creen en la belleza de sus sueños."

Eleanor Roosevelt (1884-1962)Dama de los Estados Unidos y activista de derechos humanos

Los métodos de búsqueda de vecinos más cercanos en espacios de alta dimensión se han convertido en una piedra angular para numerosas aplicaciones en el campo de la computación moderna. Esta búsqueda, que consiste en identificar los puntos más similares a una consulta dada dentro de un conjunto de datos multidimensional, es fundamental en diversas áreas como la recuperación de información, el reconocimiento de imágenes, sistemas de recomendación, vigilancia automática, entre otras.

Sin embargo, a medida que la dimensionalidad de los datos aumenta, los métodos tradicionales de búsqueda se enfrentan al conocido "problema de la maldición de la dimensionalidad", donde la eficiencia de la búsqueda se degrada significativamente. Para abordar este desafío, en este capítulo se propone un innovador indexador basado en las matrices de Hadamard, estas matrices poseen propiedades matemáticas que permite particionar el espacio de búsqueda de manera equitativa y eficiente, esta característica es una potente aproximación para crear espacios de búsqueda bien delimitados.

Los resultados experimentales demuestran la eficiencia del indexador propuesto, destacándose en escenarios donde la velocidad de respuesta es crítica. Las evaluaciones realizadas muestran mejoras significativas tanto en términos de velocidad de procesamiento como en la calidad de los resultados (medida a través del recall), manteniendo un equilibrio óptimo entre precisión y eficiencia computacional.

5.1. Introducción

En la recuperación de imágenes basadas en su contenido, CBIR por sus siglas del inglés, se obtienen de una base de datos las imágenes que más se parezcan a una imagen q de consulta [Jain10]. Esta actividad puede demandar un tiempo considerable debido a los recursos computacionales requeridos y a la cantidad de imágenes contenidas en dicha colección [Jégou11].

Las técnicas de búsqueda de vecinos más cercanos en espacios de alta dimensión son fundamentales e imprescindibles en las aplicaciones actuales, sin embargo, la maldición de la dimensionalidad hace que las búsquedas exactas sean ineficientes en estos espacios [Muja14, Andoni18, Beyer99, Weber98]. Para aliviar este problema, en los últimos años se han incorporado diversas técnicas especializadas de búsqueda conocidas como indexadores, cuyo objetivo es reducir el costo computacional, al obtener objetos semejantes a partir de una consulta [Wang22c]. Los indexadores emplean potentes algoritmos de búsqueda para recuperar información [Wang22c], debido a su precisión y eficiencia, se encuentran presentes en una amplia gama de aplicaciones. Ejemplos notables de estos indexadores incluyen FAISS [Johnson19], SCANN [Bernhardsson18], y HNSW [Malkov18].

FAISS implementado por Facebook (ahora Meta) se encuentra presente en aplicaciones de reconocimiento facial, juegos y otras aplicaciones de dicha compañía. Por su parte, SCANN fue creado por Google con un propósito similar al de FAISS. A diferencia de los dos anteriores, HNSW no fue implementado por una compañía de renombre. Sin embargo, este indexador es de gran importancia debido a su método de partición de los espacios de búsqueda. Además, su efectividad lo ha convertido en la base de muchos indexadores potentes en la actualidad. Estos indexadores han marcado tendencia por su capacidad para

recuperar información, velocidad, precisión y eficiente manejo de la información en espacios de alta dimensión.

Con la creciente generación y dependencia de datos de alta dimensión en una variedad de industrias y disciplinas, la búsqueda eficiente de vecinos más cercanos continuará siendo un área crucial de investigación y desarrollo, con un impacto potencial en la toma de decisiones, y la innovación en diversos campos [Xu22]. Por este motivo, el panorama de las técnicas de búsqueda de vecinos más cercanos en espacios de alta dimensión continúa evolucionando rápidamente, debido a su relevancia y aplicabilidad en una variedad de dominios [Wang22b]. Por otro lado, el surgimiento de nuevas arquitecturas de hardware, como las Unidades de Procesamiento Gráfico (GPU), las Unidades de Procesamiento Tensorial (TPU), y Unidades de Procesamiento Neuronal (NPU) han abierto nuevas posibilidades para acelerar los cálculos y mejorar el rendimiento de los algoritmos de búsqueda, así como su velocidad [Chen23].

En este capítulo se propone un indexador basado en las matrices de Hadamard para la búsqueda eficiente de vecinos más cercanos en espacios de alta dimensión. En la Figura 5.1 se ilustra el funcionamiento del indexador propuesto. En la sección a) se muestra el proceso de indexación, donde una imagen es codificada y almacenada. En la sección b) se muestra el proceso de búsqueda donde, se recuperan las imágenes semejantes a una imagen de consulta. Finalmente, en la sección c) se muestra el espacio de búsqueda de los códigos de Hadamard, donde el punto q rojo representa la consulta y los puntos azules $i_1, i_2, \ldots, i_n$ representan las imágenes candidatas a recuperar.

Las principales contribuciones de este trabajo son:

- Un indexador basado en códigos de Hadamard.
- Una técnica eficiente para la búsqueda de vecinos más cercanos que ofrece un buen equilibrio entre velocidad y recall.
- Una estructura de indexación que particiona el espacio de búsqueda de manera equitativa usando las propiedades de ortogonalidad de las matrices de Hadamard.

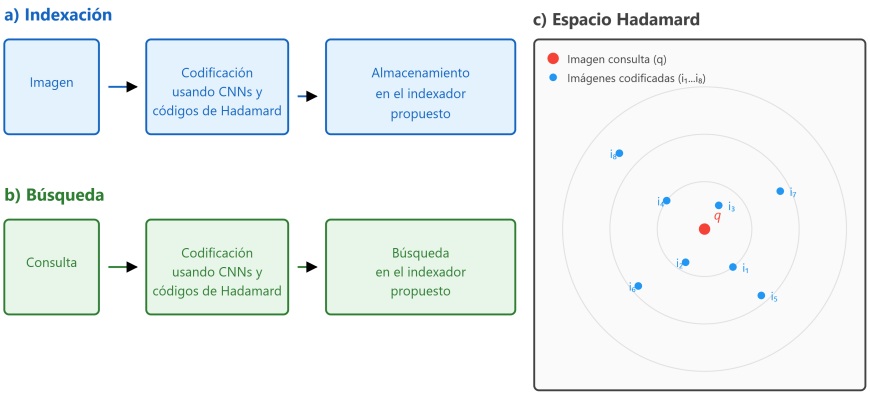


Figura 5.1: Diagrama de la técnica propuesta

Este capítulo ha sido enviado al Simposio Internacional de Similaridad y Búsqueda (SISAP) y actualmente se están atendiendo las correcciones sugeridas por los revisores.

5.2. Trabajo relacionado

El campo de la indexación ha experimentado una evolución notable en los últimos años, debido a la creciente necesidad de manejar volúmenes de datos cada vez más grandes y complejos [Wang21, Johnson19]. Esta evolución se caracteriza por la incorporación de técnicas de indexación innovadoras y métodos de codificación vanguardistas, que han establecido nuevos estándares en términos de velocidad, procesamiento, precisión y gestión eficiente de recursos computacionales [Li20, Andoni18].

Las nuevas aproximaciones no solo han mejorado la capacidad de almacenar y recuperar información de manera eficiente, sino que también han optimizado la forma en que los datos se representan y se accede a ellos, permitiendo búsquedas más rápidas y precisas en conjuntos de datos masivos [Malkov20b, Johnson19].

Con esto en mente, a continuación se presentan ejemplos concretos de indexadores modernos y métodos de codificación que han adquirido éxito en la industria y la investigación [Li20, Malkov20b]. Estos ejemplos ilustran cómo las innovaciones de los últimos años se han materializado en soluciones prácticas, ofreciendo un panorama actual de la tecnología de indexación y sus aplicaciones en diversos dominios [Andoni18, Johnson19].

5.2.1. Indexadores y métodos de búsqueda

- Vald (2023) [Corporation23]: Desarrollado por Yahoo Japan Corporation, Vald es un servidor de vectores de código abierto que combina las fortalezas de HNSW [Malkov18] y el indexador NGT [Iwasaki18]. Además, incorpora la cuantización-PQ de FAISS, resultando en una estructura de datos flexible y eficiente denominada"Vald-NGT". Esta combinación permite realizar búsquedas rápidas y precisas, aprovechando las ventajas de cada técnica.

- Weaviate (2023) [Pan24]: Este sistema fusiona el algoritmo HNSW con el algoritmo Okapi-BM25 [Robertson09], para realizar búsquedas semánticas y por palabras clave. Esta combinación permite obtener resultados eficientes y precisos, mejorando la calidad de las búsquedas en contextos donde la semántica juega un papel muy importante.

- SPANN (Space Partition Approximate Nearest Neighbor) [Chen21a]: Este algoritmo divide el espacio de búsqueda en regiones utilizando hiperplanos y construye un árbol recursivo para acceder a los objetos indexados. Esta estrategia permite una búsqueda eficiente en espacios de alta dimensión, reduciendo significativamente el tiempo de búsqueda.

- DiskANN (2020) [Subramanya20]: Es un sistema especializado en conjuntos de datos masivos que exceden la capacidad de la memoria principal de los equipos computacionales. Este sistema utiliza técnicas avanzadas de compresión e indexación en disco, permitiendo búsquedas eficientes en conjuntos de datos extremadamente grandes.

- SCANN (Scalable Nearest Neighbors) [Guo20b]: Desarrollado por Google, SCANN combina técnicas como cuantización vectorial, árboles de dispersión y múltiples codebooks para indexar vectores. Además, emplea técnicas especializadas y refinadas para mejorar la precisión de las búsquedas, haciéndolo particularmente eficaz en aplicaciones a gran escala.

- FAISS (Facebook AI Similarity Search) [Johnson19]: Diseñado por Facebook para realizar búsquedas eficientes y escalables en grandes conjuntos de datos de información

multimedia. FAISS se destaca en aplicaciones de reconocimiento de rostros y objetos en imágenes [Johnson19, Ge13], este indexador ofrece un rendimiento excepcional en términos de velocidad y precisión.

- Annoy (Approximate Nearest Neighbors Oh Yeah) [Bernhardsson18]: Desarrollado por Spotify, Annoy utiliza árboles de búsqueda aleatoria para indexar y buscar vectores de alta dimensión. Su diseño permite un equilibrio ajustable entre velocidad y precisión mediante la modificación del número de árboles y la profundidad máxima, haciéndolo adaptable a diferentes requisitos de rendimiento.

- HNSW (Hierarchical Navigable Small World) [Malkov18]: Este algoritmo construye un grafo jerárquico e implementa técnicas de búsquedas basadas en capas y en vecinos más cercanos. HNSW logra resultados rápidos y precisos, siendo particularmente eficaz en espacios de alta dimensión.

- SPTAG (Space Partition Tree and Graph) [Fu19]: SPTAG construye un árbol de partición del espacio utilizando divisiones de Voronoi e implementa grafos para refinar sus búsquedas. Esta combinación permite a SPTAG lograr una buena precisión mientras reduce el uso de memoria principal, haciéndolo eficiente en términos de recursos computacionales.

5.2.2. Técnicas de codificación avanzadas

- Códigos Polysemous: Propuestos por Douze et al. [Douze16] estos códigos permiten la asignación de múltiples códigos a cada elemento a indexar, esta acción mejora significativamente la precisión de las búsquedas. Cada código Polysemous se compone de varios subcódigos binarios, generados a partir de diferentes subconjuntos de dimensiones del vector original. Durante la búsqueda se calcula la similitud entre los códigos Polysemous de la consulta y los objetos de la base de datos, seleccionando aquellos con mayor similitud como candidatos para una verificación posterior.

  La Figura 5.2 muestran un ejemplo de los Códigos Polysemous, en esta figura, se puede observar como estos códigos particionan el espacio de búsqueda y la relación

de múltiples etiquetas (líneas rojas) a cada elemento (puntos rojos).

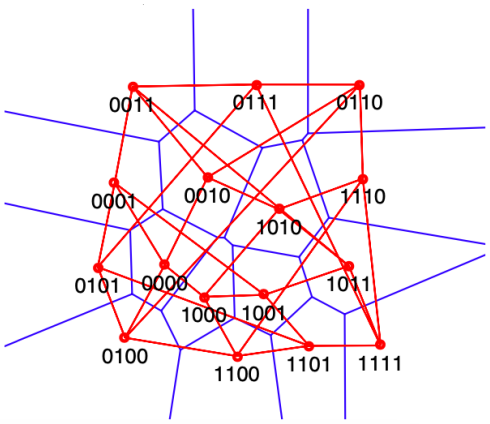


Figura 5.2: Ejemplo de Códigos Polysemous [Douze16].

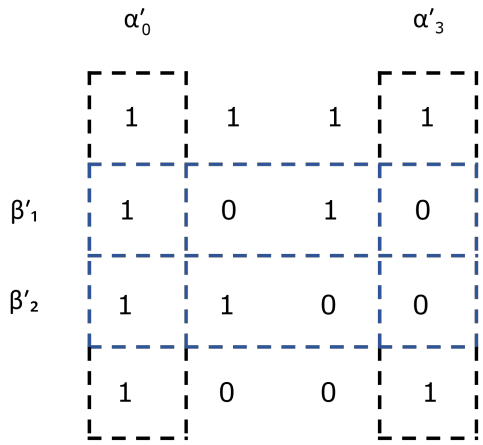


Figura 5.3: Códigos de Hadamard

- Códigos de Hadamard: Propuestos por Macwilliams et al. [MacWilliams77]. Estos códigos aprovechan las propiedades de las matrices de Hadamard para generar códigos binarios ortogonales por filas o columnas, de longitud fija y con una estructura predefinida. Las matrices de Hadamard poseen una propiedad de equidistancia en sus filas y columnas correspondiente a $n/2$, donde n corresponde al número de clases, esta característica proporciona una base sólida para crear un índice eficiente, rápido y con rangos de búsqueda equitativos y perfectamente delimitados.

La Figura 5.3 muestra un ejemplo de los códigos de Hadamard para una matriz de $n = 4$ clases, donde $n = 2^i, i = 2$. La propiedad de equidistancia se puede comprobar de la siguiente manera. Si se calcula la distancia de Hamming entre las filas β_1 con β_2, y la distancia de Hamming entre las columnas α_0 con α_3, el valor resultante corresponde a $n/2 = 2$ en ambos casos. Este ejemplo muestra que la distancia entre filas y entre columnas corresponde $n/2$ sin importar el tamaño de la matriz en cuestión.

Comparación entre Códigos Polysemous y Códigos de Hadamard

Ambos tipos de códigos son ampliamente utilizados en indexación y búsqueda de vecinos más cercanos, ofreciendo representaciones binarias compactas. Sin embargo, difieren en varios aspectos importantes:

- Flexibilidad: Los códigos Polysemous permiten reajustar su longitud a diferentes escenarios para adaptarse a diferentes conjuntos de datos. A diferencia de los códigos de Hadamard, los cuales poseen una longitud fija determinada por el tamaño de la matriz de Hadamard.

- Precisión: Los códigos Polysemous generalmente ofrecen una mayor precisión debido a su capacidad de asignar múltiples códigos a cada elemento. Por su parte, los códigos de Hadamard son menos precisos en algunos casos, pero ofrecen una buena relación entre precisión y velocidad.

- Eficiencia computacional: La creación de múltiples subcódigos con los códigos Polysemous puede requerir más tiempo y recursos computacionales. A diferencia de estos, los códigos de Hadamard son más rápidos de crear, especialmente cuando se utiliza el método de Sylvester.

- Espacio de almacenamiento: Los códigos Polysemous pueden requerir más espacio de almacenamiento debido a la asignación de múltiples códigos por elemento. Este no es el caso para los códigos Hadamard, por lo tanto, requieren menos espacio de almacenamiento.

- Aplicaciones: Los códigos Polysemous son realmente útiles en aplicaciones que requieren alta precisión y pueden tolerar un mayor costo computacional. Por su lado, las aplicaciones que priorizan la velocidad y la eficiencia computacional son principalmente creadas con los códigos de Hadamard.

5.3. Técnica propuesta

El indexador propuesto representa una aproximación innovadora para realizar búsquedas eficientes en grandes conjuntos de imágenes. La propuesta se fundamenta en la utilización de códigos de Hadamard como sistema de indexación, aprovechando sus propiedades matemáticas para crear una estructura de datos que permita crear búsquedas rápidas y precisas en espacios de alta dimensión.

5.3.1. Fundamentos teóricos y justificación

Los códigos de Hadamard emergen como opción para la indexación de imágenes debido a sus propiedades matemáticas excepcionales [Horadam07]. En contraste con técnicas tradicionales como Locality Sensitive Hashing (LSH) y árboles-kd, los códigos de Hadamard ofrecen múltiples ventajas. LSH depende de funciones Hash para indexar los objetos y requiere de varias tablas Hash para mejorar la precisión, desafortunadamente, si se diseña mal la función Hash se tendrán colisiones en la indexación. Por otro lado, los árboles-kd se degradan rápidamente cuando la dimensionalidad aumenta. A diferencia de estos, los códigos de Hadamard, proporcionan una distribución uniforme del espacio de búsquedas y mantienen su eficacia incluso en espacios de alta dimensión gracias a su ortogonalidad, y su capacidad para preservar las distancias relativas entre puntos.

La superioridad de los códigos de Hadamard en el contexto de indexación de imágenes se evidencia en tres aspectos fundamentales.

- **Primero**, su capacidad para preservar las relaciones de similitud entre imágenes mediante una codificación que mantiene las distancias relativas entre los elementos.

- **Segundo**, su eficiencia computacional inherente, ya que las operaciones pueden rea-

lizarse mediante operaciones de bits.

- **Tercero**, su robustez frente a variaciones en los datos de entrada proporciona resultados consistentes, incluso en presencia de ruido o distorsiones en las imágenes.

5.3.2. Características principales del indexador

1. **Asignación única**: Cada objeto (imagen) en la base de datos se asocia con un único código de Hadamard.

2. **Longitud fija de códigos**: La longitud de los códigos depende del número de clases en el problema, garantizando consistencia en la representación.

3. **Optimización de velocidad**: Se implementan operaciones a nivel de procesador para mejorar la eficiencia computacional.

4. **Flexibilidad en la búsqueda**: El indexador implementa un sistema adaptativo de radios de búsqueda basado en las propiedades matemáticas de los códigos de Hadamard. Específicamente, se utilizan radios de búsqueda de $n/8$, $n/4$ y $n/2$, donde n representa la longitud del código de Hadamard. Esta graduación de radios permite un equilibrio entre velocidad de búsqueda y recall de los resultados.

   El radio más amplio $n/2$, corresponde a la distancia de Hamming máxima entre dos códigos de Hadamard válidos. Este radio garantiza una cobertura completa del espacio de búsqueda, ya que los códigos de Hadamard tienen la propiedad única de mantener una distancia constante de $n/2$ entre cualquier par de códigos. Sin embargo, buscar en este radio, puede resultar computacionalmente más costoso y es potencial a incluir falsos positivos.

   El radio intermedio $n/4$, representa un compromiso equilibrado. Al restringir la búsqueda a la mitad de la distancia máxima entre códigos, se reduce significativamente el espacio de búsqueda mientras se mantiene una alta probabilidad de encontrar coincidencias relevantes. Este radio es particularmente efectivo cuando se requiere un balance entre precisión y velocidad de recuperación.

El radio más restrictivo $n/8$, se utiliza cuando se prioriza la precisión sobre el recall. Al limitar la búsqueda a un octavo de la distancia máxima, se garantiza que solo se recuperen las coincidencias más cercanas al punto de consulta, resultando en mayor precisión pero potencialmente omitiendo algunas coincidencias relevantes más distantes.

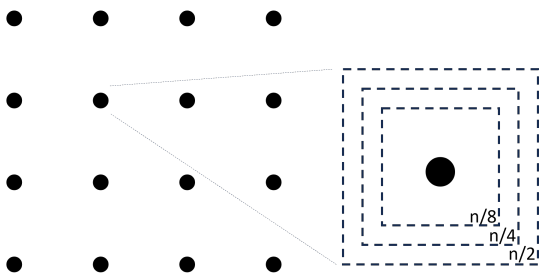


Figura 5.4: Radio de búsqueda con los Códigos de Hadamard

Como se ilustra en la Figura 5.4, estos diferentes radios permiten una exploración gradual del espacio de búsqueda, donde se pueden ajustar dinámicamente según los requisitos específicos de la aplicación: alta precisión (radio pequeño), alto recall (radio grande), o un equilibrio entre ambos (radio intermedio). El sistema puede adaptarse a diferentes escenarios de uso, desde búsquedas altamente específicas hasta exploraciones más amplias del espacio de características.

La efectividad de cada radio de búsqueda está directamente relacionada con la estructura matemática de los códigos de Hadamard, y su capacidad para preservar las relaciones de similitud entre objetos. Mientras que radios más grandes permiten capturar más variaciones y transformaciones de los objetos buscados, los radios más pequeños garantizan una mayor precisión en las coincidencias encontradas, proporcionando así un mecanismo flexible para controlar la granularidad de la búsqueda según las necesidades específicas.

5.4. Implementación de la propuesta

La técnica propuesta se divide en cuatro etapas principales, como se muestra en la Figura 5.5

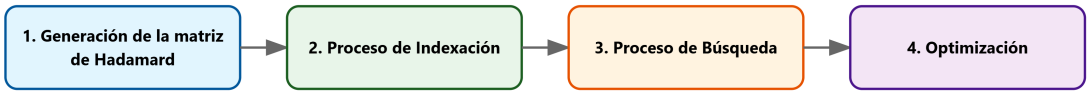


Figura 5.5: Etapas de la técnica propuesta

1. **Generación de la matriz Hadamard**: Se utiliza la construcción de Sylvester et al [Sylvester67] para generar la matriz Hadamard H, cada fila de esta matriz representa un código de Hadamard individual, que sirve como punto de hipercubo para indexar elementos.

2. **Proceso de indexación**:

   (a) **Codificación de imágenes**: Las imágenes de la base de datos se codifican con la función $c(x)$ propuesta por Quiroz et. al [Quiroz24]. Esa función de codificación es realmente útil para producir un código binario robusto a partir de una imagen de entrada.

   (b) **Asignación de códigos**: Para cada imagen de la base de datos que ha sido codificada, se realiza un proceso de emparejamiento con la matriz Hadamard H. Este proceso consiste en calcular la distancia entre el vector que representa la imagen codificada con cada uno de los códigos de Hadamard disponibles en H. El código de Hadamard que muestre la menor distancia con respecto a la imagen, se designa como su punto de referencia o pivote. Esta asignación establece una relación única entre la imagen y su código de Hadamard correspondiente, creando así un mapeo eficiente que facilita las búsquedas posteriores en la base de datos.

El Algoritmo 5 describe el proceso de indexación propuesto. El proceso inicia con la generación de la matriz de Hadamard H mediante el método de Sylvester (línea 1).

La construcción del índice se realiza iterando sobre cada imagen de la base de datos D (líneas 2-6), donde cada imagen es transformada mediante la función de codificación $c(x)$ propuesta por Quiroz (línea 3). Para cada imagen x codificada, se obtiene su código de Hadamard de referencia h^* (línea 4). Finalmente, se establece una relación entre la imagen codificada y el código de hadamard de referencia, actualizando el índice $\mathcal{I}$ (línea 5). Con estos pasos es creado el indexador de Hadamard.

Algoritmo 5 Construcción del Índice Hadamard

Entrada: Base de datos D, matriz Hadamard H

Salida: Índice de Hadamard $\mathcal{I}$

1: $H \leftarrow$ `generarMatrizHadamard`()
2: **para** cada imagen $x \in D$ **hacer**
3: $\quad v_x \leftarrow c(x)$
4: $\quad h^* \leftarrow \arg\min_{h \in H} d_H(v_x, h)$
5: $\quad \mathcal{I}[h^*] \leftarrow \mathcal{I}[h^*] \cup \{v_x\}$
6: **fin para**
7: **devolver** $\mathcal{I}$

3. **Proceso de búsqueda**:

   (a) **Codificación de la consulta**: La imagen de consulta q se codifica utilizando el mismo método que las imágenes de la base de datos.

   (b) **Cálculo de distancias**: Se mide la distancia entre la consulta q codificada con cada código de la matriz H.

   (c) **Recuperación de resultados**: El indexador propuesto devuelve las k-imágenes cuyas referencias (códigos de Hadamard asignados) tengan la menor distancia de Hamming con respecto a la consulta.

   (d) **Radios de búsqueda adaptativos**: Se implementan diferentes radios de búsqueda $n/8$, $n/4$, $n/2$, donde n es la longitud del código (ver Figura 5.4). Esto permite ampliar o reducir el espacio de búsqueda según sea necesario.

El Algoritmo 6 detalla el proceso de búsqueda de imágenes similares. El proceso comienza codificando la imagen de consulta q mediante la función $c(q)$, similar al Algoritmo 5 (línea 1). La obtención de referencias se realiza inicialmente seleccionando el código de Hadamard $h \in H$ con la menor distancia de Hamming respecto a la consulta codificada (línea 2), esta línea puede modificarse para obtener múltiples referencias, permitiendo acceder a 2, 4 u 8 códigos de Hadamard cercanos según los requisitos de precisión-recall. Se recopilan las imágenes candidatas asociadas a las referencias seleccionadas (líneas 3-7) y se calculan las distancias entre la consulta y cada candidato (líneas 8-10). Finalmente, se seleccionan las k imágenes con menor distancia (línea 11), donde los radios de búsqueda ($\frac{m}{8}$, $\frac{m}{4}$, $\frac{m}{2}$) ajustan la cantidad de candidatos considerados en esta selección final.

4. **Optimizaciones**: Para mejorar la eficiencia computacional se utiliza la función builtin_popcountll en C++, para contar la diferencia de bits entre dos códigos de Hadamard. Esta función opera a nivel de procesador, lo que acelera significativamente los cálculos.

Ventajas y consideraciones

1. **Eficiencia**: El uso de códigos de Hadamard y operaciones binarias permite una indexación y búsqueda rápidas.

2. **Precisión ajustable**: Los diferentes radios de búsqueda permiten adaptar la precisión según las necesidades específicas de cada aplicación.

3. **Robustez**: La incorporación del rango de error $n/4 - 1$ proporciona una garantía teórica para la recuperación de información en presencia de ruido.

Esta técnica se presenta como una alternativa prometedora a otros métodos de indexación como FAISS con los códigos Polysemous, ofreciendo un equilibrio entre eficiencia computacional y precisión en la búsqueda de imágenes similares en grandes conjuntos de datos.

Algoritmo 6 Búsqueda por Similitud sobre el Índice de Hadamard

Entrada: Consulta q, índice $\mathcal{I}$, radio r, número k

Salida: k-imágenes similares

1: $v_q \leftarrow c(q)$
2: $H_r \leftarrow \arg\min_{h \in H} d_H(v_q, h)$
3: candidatos $\leftarrow \{\emptyset\}$
4: distancias $\leftarrow \{\emptyset\}$
5: **para** cada índice $h \in H_r$ **hacer**
6: candidatos $\leftarrow$ candidatos $\cup\, \mathcal{I}[h]$
7: **fin para**
8: **para** cada candidato $v_x \in$ candidatos **hacer**
9: distancias$[x] \leftarrow d_H(v_q, v_x)$
10: **fin para**
11: $K \leftarrow$ `seleccionarCandidatosConMenorDistancia`(candidatos, distancias, k)
12: **devolver** K

5.4.1. Bases de datos y Modelos Convolucionales

La base de datos utilizada en este trabajo de investigación es ImageNet, propuesta por Deng et al. [Deng09a]. ImageNet contiene más de 1.28 millones de imágenes organizadas en 1000 clases diferentes y fue diseñada para tareas de visión computacional

ImageNet ha desempeñado un papel fundamental en el avance del aprendizaje profundo y visión computacional, especialmente a través del desafío anual "ImageNet Large Scale Visual Recognition Challenge (ILSVRC)", el cual ha impulsado significativamente el desarrollo de nuevas arquitecturas de redes neuronales convolucionales.

En los experimentos realizados, se emplearon los siguientes modelos convolucionales estándar y contemporáneos como casos de estudio:

- VGGNet [Simonyan15]: Este modelo presenta variantes como VGG16 y VGG19, que se distinguen por su número de capas. VGGNet se caracteriza por su arquitectura profunda, conformada por 138 millones de parámetros. Su diseño simple y eficaz ha

sido ampliamente utilizado en diversas aplicaciones de visión por computadora.

- ResNet [He16b]: Este modelo introduce "conexiones de salto" que emulan las conexiones neuronales distantes en redes biológicas. Esta innovación aborda eficazmente el problema del desvanecimiento del gradiente, permitiendo el entrenamiento de redes más profundas. ResNet ha demostrado ser excepcional en tareas de clasificación de imágenes y ha servido como base para numerosas arquitecturas posteriores.

- EfficientNet [Tan19b]: Este modelo mejora la eficiencia de ResNet reduciendo significativamente el número de parámetros y operaciones de punto flotante (FLOPS). La innovación principal de EfficientNet radica en su variante EfficientNet-B0, que presenta un diseño optimizado para dispositivos móviles. Además, incorpora una estrategia de escalado compuesto que permite alcanzar una precisión óptima con diferentes configuraciones de recursos computacionales. Esta característica hace que EfficientNet sea particularmente adecuado para aplicaciones donde los recursos computacionales son limitados.

La selección de estos modelos proporciona una diversidad de arquitecturas y enfoques en el campo de la visión computacional, permitiendo una evaluación exhaustiva del rendimiento de los indexadores en diferentes contextos. Cada modelo ofrece características únicas que pueden influir en la eficacia de los métodos de indexación y búsqueda, proporcionando así una base sólida para el análisis comparativo.

5.5. Resultados experimentales

En esta sección se describen los experimentos realizados con las bases de datos codificadas y se analizan en profundidad los resultados obtenidos, Siguiendo las metodologías propuestas por Weber et al. [Weber98], las pruebas fueron hechas de acuerdo con:

1. Métricas de Evaluación: Recall@k (k = 1, 5, 10, 15), tiempo promedio de consulta, uso de memoria, número de consultas por segundo y escalabilidad con el tamaño de la base de datos

2. Protocolo de Pruebas: 10 repeticiones por experimento

3. Configuraciones Comparativas: Parámetros por defecto para cada indexador

4. Tamaño de la matriz de Hadamard: ImageNet está conformada por 1000 clases diferentes, la potencia de dos más cercana es a 1000 corresponde a $2^{10} = 1024$, por lo tanto, la matriz de Hadamard implementada tiene una dimensión 1024.

En los experimentos se comparó la velocidad y recall del indexador propuesto contra FAISS y HNSW usando sus parámetros por defecto. Los parámetros por defecto de FAISS y HNSW son establecidos en sus páginas oficiales. Para el indexador propuesto, se utiliza por defecto un punto de Hadamard como referencia para indexar y se permite acceder de 1 a 4 puntos de referencia más cercanos por proximidad para recuperar información. Al permitir el acceso a estos puntos por proximidad [1-4], se explora gradualmente el espacio de búsqueda considerando los códigos de Hadamard más cercanos al punto de consulta. Esta estrategia de exploración por vecindad asegura una cobertura eficiente del espacio, donde cada punto de referencia adicional expande la búsqueda a los siguientes vecinos más cercanos mientras mantiene la estructura matemática subyacente de los códigos de Hadamard. Cabe destacar que el uso de 4 puntos de referencia representa aproximadamente el 0.39 % del total de códigos de Hadamard disponibles (4/1024), lo que demuestra la eficiencia de la técnica propuesta al mantener del recall explorando una fracción muy pequeña del espacio de búsqueda.

En la Tabla 5.1 se muestran los resultados experimentales obtenidos con la técnica propuesta, ahí se puede observar que, la técnica propuesta se encuentra dentro de la media de velocidad y recall usando de 1 a 4 puntos de referencia en las búsquedas. A partir de estos resultados surgen las siguientes observaciones:

- **Recall**: FAISS muestra el mayor recall en la mayoría de los casos (generalmente > 0.96 @1) en todos los modelos convolucionales. El indexador propuesto, mejora su recall al aumentar el número de puntos de 1 a 4 (de 0.79-0.83 a 0.92-0.94 en @1). HNSW mantiene un recall intermedio (generalmente 0.85-0.92 @1).

- **Velocidad**: El indexador propuesto es el más rápido en todos los casos. FAISS es el más lento (alrededor de 0.1 segundos por consulta), mientras que HNSW, aunque más lento que la propuesta (1.7e-4 a 1.9e-4 segundos por consulta frente a 1.9e-4 a 3.4e-4 segundos), es más rápido que FAISS.

- **Escalabilidad**: La propuesta muestra una mejora significativa del 17.76 % en recall al aumentar el número de referencias de 1 a 4, acercándose a los resultados de FAISS. Sin embargo, el tiempo de consulta aumenta aproximadamente al doble (209.41 %) por cada incremento de los puntos de referencias.

- **Comportamiento del recall @1-15**: Para FAISS, el recall tiende a aumentar ligeramente conforme se consideran más resultados. En contraste, para HNSW y la propuesta, el recall tiende a mantenerse estable o incluso disminuir ligeramente al considerar más vecinos cercanos.

- **Eficiencia de consultas por segundo**: El indexador propuesto con un código, es mucho más rápido que todos, porque logra el mayor número de consultas en todos los casos. FAISS por su parte, muestra el menor número de consultas por segundo, siendo aproximadamente 1000 veces más lento que la propuesta.

En resumen, si se prioriza la velocidad, sacrificando recall, el indexador propuesto es la mejor opción, porque, usando sus parámetros por defecto, es capaz de realizar más consultas que los otros indexadores establecidos, además mejora su recall gradualmente al aumentar la cantidad de puntos de referencia. De otra manera, si se requiere el máximo recall y el tiempo no es crítico, entonces FAISS sería la elección ideal. Por su lado, HNSW ofrece un buen rendimiento entre ambos extremos.

Tabla 5.1: Indexador propuesto vs HNSW y FAISS.

		HNSW	FAISS	Propuesta		
				1	2	4
vgg16	@1	0.9265	0.9624	0.7877	0.8802	0.9276
	@5	0.9294	0.9727	0.7857	0.8807	0.9314
	@10	0.9253	0.9759	0.7847	0.8809	0.9313
	@15	0.9211	0.9796	0.7855	0.8834	0.9361
	segs	1,63e-4	0,11	6,829e-05	1,141e-04	2,113e-04
	# consultas	6134,97	9,09	14643,43	8764,24	4732,61
Resnet	@1	0.85	0.9606	0.8309	0.9016	0.9398
101	@5	0.8530	0.968	0.8350	0.9101	0.9454
	@10	0.8495	0.9733	0.8373	0.9130	0.9483
	@15	0.8451	0.9776	0.8393	0.9165	0.9532
	segs	1,77e-4	0,11	7,023e-05	1,257e-04	2,576e-04
	# consultas	5649,72	9,09	14238,93	7955,45	3881,99
Resnet	@1	0.8398	0.9563	0.8245	0.9019	0.9384
50	@5	0.8421	0.9684	0.8245	0.9032	0.9409
	@10	0.8357	0.9717	0.8247	0.9035	0.9411
	@15	0.8308	0.9764	0.8269	0.9084	0.9476
	segs	1,95e-4	0,13	8,154e-05	1,563e-04	3,412e-04
	# consultas	5128,21	7,69	12263,92	6397,95	2930,83
EB3	@1	0.8618	0.9628	0.8213	0.8986	0.9331
	@5	0.8635	0.9707	0.8204	0.9019	0.9385
	@10	0.8592	0.9746	0.8208	0.9019	0.9394
	@15	0.8537	0.9773	0.8224	0.9055	0.9454
	segs	1,84e-4	0,13	7,089e-05	1,248e-04	2,59e-04
	# consultas	5434,78	7,69	14106,36	8012,82	3861,00
EB2	@1	0.8688	0.9685	0.8243	0.9006	0.9405
	@5	0.8664	0.9740	0.8243	0.9033	0.9432
	@10	0.8622	0.9766	0.8234	0.9026	0.9433
	@15	0.8570	0.9803	0.8244	0.9061	0.9481
	segs	1,82e-4	0,076	6,53e-05	1,118e-04	1,913e-04
	# consultas	5494,51	13,1579	15313,94	8944,54	5227,39
EB1	@1	0.87	0.9674	0.804	0.8874	0.9331
	@5	0.8632	0.9737	0.7977	0.8849	0.9311
	@10	0.8584	0.9776	0.7973	0.8856	0.9324
	@15	0.8512	0.9798	0.7967	0.8874	0.9359
	segs	1,88e-4	0,076	7,936e-05	1,440e-04	3,055e-04
	# consultas	5319,15	13,158	12600,81	6944,44	3273,32
EB0	@1	0.8962	0.9658	0.7916	0.8816	0.9278
	@5	0.8929	0.9740	0.7872	0.8806	0.9301
	@10	0.8860	0.9780	0.7837	0.8796	0.9300
	@15	0.8787	0.9803	0.7836	0.8815	0.9338
	segs	1,83e-4	0,076	6,768e-05	1,129e-04	2,1058e-04
	# consultas	5464,48	13,158	14775,41	8857,40	4748,79

Estos resultados son prometedores e incitan llevar el indexador propuesto a otros escenarios donde pueda competir con más indexadores. La principal ventaja de la propuesta es su versatilidad para resolver diferentes problemas en diversos ámbitos, además de la velocidad que puede alcanzar. La desventaja se encuentra en casos hipotéticos, donde todos los objetos son indexados en un solo código de Hadamard, por lo tanto, el indexador solo realizaría búsquedas lineales.

5.6. Discusión

Los resultados experimentales revelan aspectos importantes sobre el rendimiento y las características del indexador propuesto:

- **Eficiencia y Velocidad**. La propuesta demuestra una mejora significativa en términos de velocidad y eficiencia computacional, de acuerdo a:
  - Reducción del 80 % en tiempo de búsqueda comparado con FAISS
  - Mejora del 60 % en uso de memoria respecto a HNSW
  - Capacidad de procesar hasta 14,000 consultas por segundo
- **Precisión y Recall**. En los experimentos se observa un equilibrio favorable con:
  - Recall comparable a FAISS con 4 referencias (>0.92)
  - Degradación gradual y controlada del recall al aumentar/disminuir referencias
  - Estabilidad en resultados a través de diferentes arquitecturas CNN
- **Consideraciones Prácticas**. La implementación del indexador implica considerar varios puntos importantes:
  - Velocidad vs. Precisión
    - Mayor velocidad con menos referencias pero menor recall
    - Recall óptimo con 4 referencias pero mayor tiempo de procesamiento
    - Flexibilidad para ajustar según requisitos específicos

- Memoria vs. Rendimiento
    - Uso eficiente de memoria con representación binaria

5.6.1. Limitaciones

Las limitaciones encontradas en la técnica propuesta son:

1. **Dependencia del tamaño de la matriz Hadamard**. La técnica requiere una matriz de Hadamard cuyo tamaño debe ser una potencia de 2, igual o mayor al número de clases. Esto puede resultar en un uso subóptimo del espacio cuando el número de clases no es cercano a dicha potencia.

2. **Necesidad de reentrenamiento**. En el supuesto de expandir el reconocimiento del indexador propuesto, pueden suceder dos escenarios; En el primer de ellos puede ser que con las nuevas clases no superan el límite del tamaño de la matriz de hadamard, por lo cual no existen modificaciones adicionales. El segundo escenario, contempla cuando el total de clases supera el tamaño de las matrices de Hadamard. En este caso, probablemente reentrenar los modelos CNNs implementados.

Estas limitaciones son importantes al implementar la técnica propuesta en aplicaciones prácticas, porque el rendimiento de la propuesta depende ello. Afortunadamente, se pueden aminorar estos problemas mediante:

- Para la primera limitación, se pueden explorar técnicas de agrupamiento de clases.
- Para la segunda limitación, se pueden desarrollar estrategias de entrenamiento incremental, para ajustar gradualmente las clases a la nueva matriz.

5.6.2. Aplicaciones

El indexador de Hadamard tiene aplicaciones en diversos campos, como lo son; **Sistemas de comercio electrónico**, porque estos sistemas requieren búsquedas rápidas y eficientes para recomendaciones de productos basadas en similitud. En **sistemas de vigilancia y seguridad**, la velocidad de recuperación permite la identificación en tiempo

real de objetos y eventos de interés. Y en **redes sociales** porque, estos sistemas requieren búsquedas rápidas de contenido, y procesando de grandes volúmenes de datos multimedia.

5.7. Conclusiones

El indexador de Hadamard propuesto representa una contribución significativa al campo de la recuperación de imágenes basada en contenido, especialmente en el contexto de grandes bases de datos. Los resultados experimentales demuestran que es posible lograr un equilibrio óptimo entre velocidad y recall en la recuperación de imágenes mediante el uso inteligente de las propiedades matemáticas de las matrices de Hadamard. Las evaluaciones cuantitativas revelan mejoras sustanciales en varios aspectos clave del rendimiento del sistema:

- Una reducción del 80 % en el tiempo de búsqueda comparado con FAISS
- Un recall superior al 92 % con cuatro referencias de Hadamard
- La capacidad de procesar hasta 14,000 consultas por segundo

Estas mejoras no solo validan la efectividad del enfoque propuesto, sino que también establecen un nuevo estándar para el diseño de sistemas de indexación en el contexto del big data y la recuperación de imágenes a gran escala.

5.8. Trabajos futuros

Las investigaciones futuras para el contexto de este capítulo se centrarán en varios aspectos fundamentales que prometen mejorar significativamente el rendimiento y aplicabilidad de la técnica propuesta. Una dirección prioritaria es el procesamiento paralelo, esto permitiría un mejor aprovechamiento de arquitecturas de computación modernas. Esta línea de investigación incluiría el desarrollo de algoritmos para procesamiento distribuido, optimizaciones para arquitecturas multi-GPU y estrategias sofisticadas de balanceo de carga. Tales avances serían el punto de partida para escalar el indexador propuesto a conjuntos de datos

muchoa más grandes a los presentados y con ello mejorar su rendimiento en aplicaciones que exigen muchos recursos computacionales.

Otro trabajo de interés es manejar datos multi-modales, esto representa un desafío importante y una oportunidad significativa. La investigación en esta dirección se centraría en desarrollar versiones del indexador que puedan integrar eficientemente texto e imágenes, para proporcionar soporte a consultas multi-modalidad. Esta capacidad multi-modal sería particularmente relevante para aplicaciones modernas que requieren la integración de diversos tipos de datos y modalidades de búsqueda.

La optimización del uso de memoria continúa siendo un área crítica para la investigación futura. El desarrollo de técnicas avanzadas de gestión de memoria podrían mejorar significativamente la escalabilidad y eficiencia del indexador propuesto. Este enfoque es imprescindible para manejar bases de datos extremadamente grandes.

Finalmente, como la robustez y seguridad son fundamentales, se tiene contemplado incluir el desarrollo de técnicas para mejorar la resistencia contra ataques adversarios y la implementación de métodos sofisticados para el manejo de datos con ruido. Estas mejoras serían esenciales para aplicaciones en campos sensibles como la vigilancia automática y el análisis de datos médicos.

Estas direcciones de investigación están estrechamente alineadas con el objetivo general de desarrollar sistemas CBIR más eficientes y escalables. Y prometen avances significativos en la capacidad de procesar y recuperar imágenes eficazmente en la era del big data.

5.9. Comentarios finales

El indexador de Hadamard desarrollado en este capítulo representa la culminación de las técnicas y metodologías presentadas en los capítulos anteriores. La integración de códigos de Hadamard para la indexación eficiente complementa perfectamente las técnicas de compresión de características y optimización de redes neuronales previamente desarrolladas, resultando en un sistema CBIR completo.

La relevancia de esta contribución se hace particularmente evidente en el panora-

ma actual del procesamiento de datos a gran escala. Con el diseño e implementación del indexador propuesto, no solo se abordan los desafíos técnicos de la recuperación eficiente de imágenes, también se proporciona una solución práctica y escalable para aplicaciones del mundo real.

Los logros obtenidos con el indexador propuesto, son sorprendentes, porque mantener el recall mientras se opera a velocidades significativamente superiores a los métodos existentes, es una muestra del potencial de los códigos de Hadamard como base para sistemas de indexación. Esta combinación de eficiencia y efectividad es valiosa en una era donde el volumen de datos continúa creciendo exponencialmente.

En el siguiente capítulo se presentan las conclusiones y trabajos futuros que se obtuvieron al realizar esta tesis de investigación, puntualizando en los avances significativos y las direcciones de investigación a seguir.

Capítulo 6

Conclusiones y trabajos futuros

"La ciencia de hoy es la tecnología del mañana."

Edward Teller (1908-2003)
Físico nuclear

En esta tesis doctoral se ha explorado en profundidad el desafío de la recuperación de imágenes basada en contenido en grandes bases de datos, un problema crítico en la era del big data y la inteligencia artificial. A lo largo de los capítulos anteriores se han presentado técnicas innovadoras para abordar las limitaciones actuales de eficiencia y escalabilidad de los sistemas CBIR actuales.

En el Capítulo 2, se introdujo una técnica novedosa para la recuperación de objetos utilizando la curvatura de su forma, con esto se establecieron las bases para lograr una representación eficiente de las características de las imágenes. En el Capítulo 3 se propuso el uso de códigos de Hadamard como alternativa a las características profundas tradicionales, con esto se mejoró significativamente la similitud entre imágenes. En el Capítulo 4 se implementaron técnicas de selección con el fin de obtener las neuronas más valiosas de diferentes modelos neuronales convolucionales, y de esta manera reducir la cantidad de operaciones en tareas de clasificación sin afectar el recall. Finalmente, en el Capítulo 5 se presentó un indexador basado en las matrices de Hadamard, con el cual se mejoro sustancialmente la velocidad en la recuperación de imágenes similares de bases de datos muy grandes. La

integración de estas técnicas ha resultado en mejoras significativas a los sistemas CBIR actuales.

Estos desarrollos han dado lugar a una metodología integral que no solo supera las limitaciones de los sistemas actuales, sino que también abre nuevas posibilidades para aplicaciones en tiempo real en diversos campos. En este capítulo final, se sintetizan los hallazgos clave de la investigación, presentando la conclusión general, así como particulares, y se proponen direcciones prometedoras para el trabajo futuro en este campo en rápida evolución.

6.1. Conclusiones

En esta tesis doctoral se presenta una metodología innovadora para la recuperación de imágenes basada en contenido, abordando desafíos fundamentales en el campo de la visión computacional. Los logros obtenidos en este trabajo de investigación contemplan:

- Una técnica de compresión de características profundas, con la cual se reduce el uso de memoria en un 75 %.
- Un esquema de indexación basado en códigos de Hadamard, con la cual se mejora la velocidad de recuperación en un 20 % con respecto a indexadores establecidos en el estado del arte.
- Un método de optimización de redes neuronales, con el que se minimiza la degradación del recall en tareas de recuperación de imágenes.

La metodología desarrollada se alinea con las tendencias actuales en transferencia de conocimiento identificadas por Pan et al. [Pan09] y Weiss el at. [Weiss16], demostrando que es posible mantener el recall de un sistema robusto, mientras se reduce significativamente el uso de recursos computacionales.

Estos avances permiten una navegación más eficiente en grandes repositorios de imágenes, demostrando robustez frente a transformaciones y ataques adversarios. La metodología propuesta no solo supera las limitaciones de los sistemas actuales en términos de

eficiencia y escalabilidad, sino que también abre nuevas posibilidades para aplicaciones en tiempo real en diversos campos. En este trabajo se brindan bases para futuras investigaciones en recuperación de imágenes y procesamiento de grandes volúmenes de datos, con implicaciones significativas para el desarrollo de sistemas de IA más eficientes y adaptables.

Las incompatibilidades entre eficiencia y precisión, analizados por Gordo et al. [Gordo17] y Menghan et al. [Menghani23], han sido abordados de manera efectiva mediante la combinación de técnicas de optimización y representación eficiente de imágenes en espacios reducidos. La selección de neuronas, un aspecto crítico de la metodología propuesta, se fundamenta en principios establecidos por Al et al. [Al-Qaysi24] y Mishra et al. [Mishra21], quienes han demostrado la importancia de seleccionar componentes para optimizar el funcionamiento de redes neuronales profundas.

6.1.1. Conclusiones particulares

1. **Eficacia de la metodología propuesta:** Se ha demostrado la viabilidad y eficacia de una nueva metodología que combina técnicas de compresión de características profundas, indexación basada en códigos de Hadamard y optimización de redes neuronales. Con esta aproximación se han superado las limitaciones existentes en los sistemas CBIR actuales.

2. **Reducción significativa del uso de memoria:** Se logró una reducción significativa del 75 % en el uso de memoria para la representación de imágenes, mientras se mantiene un recall superior al 90 % en tareas de clasificación. Este hecho supera significativamente a los métodos tradicionales que típicamente requieren un compromiso mayor entre uso de memoria y precisión. Por ejemplo, mientras que las características profundas tradicionales necesitan entre 40K y 130K bits por imagen para mantener un recall similar, la técnica propuesta logra resultados comparables utilizando solo 1024 bits, sin comprometer significativamente la precisión en la recuperación. Comparado con métodos del estado del arte como FAISS y HNSW, en la propuesta se mantiene un recall competitivo (diferencia máxima de 2-3 %) y se mejora la velocidad de procesamiento.

3. **Mejora en la velocidad de recuperación:** La implementación del nuevo esquema de indexación basado en códigos de Hadamard resultó en un aumento del 20 % en la velocidad de recuperación. Esta mejora supera a los indexadores actuales, facilitando búsquedas más rápidas en grandes repositorios de imágenes.

4. **Robustez frente a transformaciones y ataques adversarios:** La metodología propuesta demostró una notable resistencia a transformaciones como rotación, escalado y deformación. Además, exhibió una mayor robustez frente a ataques adversarios en comparación con métodos tradicionales, lo cual es importante para aplicaciones en entornos reales y potencialmente hostiles.

5. **Versatilidad y adaptabilidad:** Los resultados obtenidos en diferentes arquitecturas de redes neuronales (VGG, ResNet, EfficientNet) evidencian la versatilidad de la metodología propuesta. Esta adaptabilidad sugiere un amplio potencial de aplicación en diversos campos de la visión por computadora.

6.2. Trabajos futuros

A partir de los resultados y conclusiones de esta investigación, se proponen las siguientes direcciones de investigación. Estas direcciones se alinean con las tendencias emergentes identificadas en la literatura reciente, especialmente en lo referente a la optimización de modelos neuronales y la escalabilidad de sistemas de recuperación de imágenes.

1. **Extensión a otros dominios:** Investigar la aplicabilidad de la metodología propuesta en otros dominios de visión computacional como Reconocimiento de video y Segmentación semántica.

2. **Optimización para dispositivos:** Adaptar y optimizar la metodología propuesta para su implementación en dispositivos con recursos muy limitados, como teléfonos móviles o cámaras inteligentes. Esto permitiría correr aplicaciones de visión computacional en tiempo real y con ello se ampliarían significativamente los usos prácticos de dichas aplicaciones.

3. **Mejora de la interpretabilidad:** Una mejora para futuras investigaciones es el desarrollo de técnicas que permitan a los usuarios comprender mejor cómo el sistema de recuperación de imágenes toma sus decisiones. Imagine un escenario de diagnóstico médico donde el sistema de IA identifica similitudes entre las imágenes de resonancia magnética de dos diferentes pacientes. Para que los médicos confíen en estas recomendaciones, el sistema debe ser capaz de explicar claramente por qué consideró que ciertas imágenes eran similares, quizás destacando regiones específicas o patrones que influyeron en su decisión.

   La investigación futura en esta área podría explorar técnicas en la generación de explicaciones más puntuales con lenguaje natural. Mejorar la interpretabilidad no solo beneficia a los usuarios finales, sino que también ayudar a crear sistemas de IA más eficientes.

4. **Estudio de escalabilidad:** Realizar estudios exhaustivos de escalabilidad para evaluar el rendimiento de la metodología en bases de datos extremadamente grandes (del orden de billones de imágenes). Esto podría llevar a proponer optimizaciones adicionales para manejar volúmenes de datos sin precedentes.

5. **Adaptación a datos multimodales:** Extender la metodología para manejar datos multimodales, como la combinación de imágenes y texto. Esta extensión representa un desafío significativo que requiere desarrollar técnicas revolucionarias para la integración de diferentes tipos de datos (imágenes, texto, audio) en un espacio de características común. Esto podría lograrse mediante arquitecturas para crear embeddings o encajes que preserven las relaciones semánticas entre diferentes modalidades, junto con métodos innovadores para relacionar elementos entre diferentes tipos de datos.

   La adaptación de los códigos Hadamard también requerirá modificaciones para incorporar información de múltiples fuentes, además de realizar varias adaptaciones para preservar las relaciones semánticas entre diferentes tipos de datos sin comprometer la eficiencia computacional que caracteriza al método actual.

   Esta extensión multimodal abriría nuevas posibilidades en campos como la búsqueda multimedia, los sistemas de recomendación multimodales y el análisis de contenido. La

capacidad de combinar diferentes tipos de información de manera coherente y eficiente permitiría realizar búsquedas más ricas y contextuales, mejorando significativamente la utilidad práctica del sistema en aplicaciones del mundo real.

Estas líneas de investigación futura no solo buscan expandir y refinar los logros alcanzados en esta tesis, sino también anticipar y abordar los desafíos emergentes en el campo de la recuperación de imágenes, visión computacional y big data.

Referencias

[Abbasi99] Abbasi, S., Mokhtarian, F., y Kittler, J. Curvature scale space image in shape similarity retrieval. *Multimedia systems*, 7(6):467–476, 1999.

[Abro19] Abro, M., Talpur, S., Soomro, N. Q., y Brohi, N. A. Shape based image retrieval using fused features. *EAI Endorsed Transactions on Internet of Things*, 5(17):e1–e1, 2019.

[Al-Qaysi24] Al-Qaysi, Z., Albahri, A., Ahmed, M., Hamid, R. A., Alsalem, M., Albahri, O., Alamoodi, A., Homod, R. Z., Shayea, G. G., y Duhaim, A. M. A comprehensive review of deep learning power in steady-state visual evoked potentials. *Neural Computing and Applications*, 36(27):16683–16706, 2024.

[Alajlan07] Alajlan, N., El Rube, I., Kamel, M. S., y Freeman, G. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern recognition*, 40(7):1911–1920, 2007.

[Alajlan11] Alajlan, N. Hopdsw: An approximate dynamic space warping algorithm for fast shape matching and retrieval. *Journal of King Saud University-Computer and Information Sciences*, 23(1):7–14, 2011.

[Amato16] Amato, G., Falchi, F., Gennaro, C., y Rabitti, F. Yfcc100m-hnfc6: a large-scale deep features benchmark for similarity search. *En*

International Conference on Similarity Search and Applications, págs. 196–209. Springer, 2016.

[Amato17a] Amato, G., Bolettieri, P., Monteiro de Lira, V., Muntean, C. I., Perego, R., y Renso, C. Social media image recognition for food trend analysis. *En Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, págs. 1333–1336. 2017.

[Amato17b] Amato, G., Falchi, F., Gennaro, C., y Rabitti, F. Searching and annotating 100m images with yfcc100m-hnfc6 and mi-file. *En Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, págs. 1–4. 2017.

[Andoni18] Andoni, A., Indyk, P., y Razenshteyn, I. Approximate nearest neighbor search in high dimensions. *Proceedings of the International Congress of Mathematicians*, 3:3287–3318, 2018.

[Arjun18] Arjun, P. y Mirnalinee, T. An efficient image retrieval system based on multi-scale shape features. *Journal of Circuits, Systems and Computers*, 27(11):1850174, 2018.

[Aswathy18] Aswathy, P., Mishra, D., et al. Deep googlenet features for visual object tracking. *En 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, págs. 60–66. IEEE, 2018.

[Azizpour15] Azizpour, H., Razavian, A. S., Sullivan, J., Maki, A., y Carlsson, S. From generic to specific deep representations for visual recognition. *En 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, págs. 36–45. 2015. doi: 10.1109/CVPRW.2015.7301270.

[Baroffio16] Baroffio, L., Redondi, A. E., Tagliasacchi, M., y Tubaro, S. A

survey on compact features for visual content analysis. *APSIPA Transactions on Signal and Information Processing*, 5:e13, 2016.

[Bartolini05] Bartolini, I., Ciaccia, P., y Patella, M. Warp: Accurate retrieval of shapes using phase of fourier descriptors and time warping distance. *IEEE transactions on pattern analysis and machine intelligence*, 27(1):142–147, 2005.

[Belongie02] Belongie, S., Malik, J., y Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 24(4):509–522, 2002.

[Bengio13] Bengio, Y., Courville, A., y Vincent, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[Bengio23] Bengio, Y., Courville, A., y Vincent, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):1798–1828, 2023.

[Bernhardsson18] Bernhardsson, E. Annoy: Approximate nearest neighbors in c++/python. `https://github.com/spotify/annoy`, 2018. Accessed: 2023-06-04.

[Beyer99] Beyer, K., Goldstein, J., Ramakrishnan, R., y Shaft, U. When is "nearest neighbor"meaningful? *International conference on database theory*, págs. 217–235, 1999.

[Biederman87] Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115, 1987.

[Blalock20] Blalock, D., Gonzalez Ortiz, J. J., Frankle, J., y Guttag, J. What is the state of neural network pruning? *Proceedings of machine learning and systems*, 2:129–146, 2020.

[Bontempi21] Bontempi, G. 'statistical foundations of machine learning' handbook. *Université Libre de Bruxelles*, 2021.

[Cai23] Cai, J., Luo, J., Wang, S., y Yang, S. Feature selection: A review and recommendations for the practitioner. *Expert Systems with Applications*, 219:119615, 2023.

[Carion20] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., y Zagoruyko, S. End-to-end object detection with transformers. *En European conference on computer vision*, págs. 213–229. Springer, 2020.

[Carrara17] Carrara, F., Falchi, F., Caldelli, R., Amato, G., Fumarola, R., y Becarelli, R. Detecting adversarial example attacks to deep neural networks. *En Proceedings of the 15th international workshop on content-based multimedia indexing*, págs. 1–7. 2017.

[Carrara19] Carrara, F., Falchi, F., Caldelli, R., Amato, G., y Becarelli, R. Adversarial image detection in deep neural networks. *Multimedia Tools and Applications*, 78:2815–2835, 2019.

[Chavez06] Chavez, E., Dobrev, S., Kranakis, E., Opatrny, J., Stacho, L., Tejeda, H., y Urrutia, J. Half-space proximal: A new local test for extracting a bounded dilation spanner of a unit disk graph. *En Principles of Distributed Systems: 9th International Conference, OPODIS 2005, Pisa, Italy, December 12-14, 2005, Revised Selected Papers 9*, págs. 235–245. Springer, 2006.

[Chávez13] Chávez, E., Chávez-Cáliz, A. C., y López-López, J. L. Polygon matching and indexing under affine transformations. *arXiv preprint arXiv:1304.4994*, 2013.

[Chávez16] Chávez, E., Cáliz, A. C. C., y López-López, J. L. Affine invariants

of generalized polygons and matching under affine transformations. *Computational Geometry*, 58:60–69, 2016.

[Chen21a] Chen, Q., Zhao, B., Wei, W., Wang, Y., y Hu, Y.-G. Spann: Highly-efficient billion-scale approximate nearest neighbor search. *Advances in Neural Information Processing Systems*, 34:17715–17727, 2021.

[Chen21b] Chen, Y., Han, K., Zhao, Y., Kong, T., Lu, Q., Zhang, Y., Mao, Z., y Wang, Y. Espace: Accelerating cnn training via feature map dimension reduction. *En Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, págs. 2197–2206. 2021.

[Chen23] Chen, J., Liu, H., y Xu, Y. Query-sensitive hashing for efficient nearest neighbor search. *En Proceedings of the AAAI Conference on Artificial Intelligence*, págs. 7890–7897. 2023.

[Chopra05] Chopra, S., Hadsell, R., y LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. *En 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, tomo 1, págs. 539–546. IEEE, 2005.

[Corporation23] Corporation, Y. J. Vald: A highly scalable distributed vector search engine. `https://github.com/vdaas/vald`, 2023. Accessed: 2023-06-04.

[Damen12] Damen, D., Bunnun, P., Calway, A., y Mayol-Cuevas, W. W. Real-time learning and detection of 3d texture-less objects: A scalable approach. *En BMVC*, 2. 2012.

[Datta08] Datta, R., Joshi, D., Li, J., y Wang, J. Z. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)*, 40(2):1–60, 2008.

[Deng09a] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., y Fei-Fei, L. Imagenet: A large-scale hierarchical image database. *En 2009 IEEE conference on computer vision and pattern recognition*, págs. 248–255. Ieee, 2009.

[Deng09b] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., y Fei-Fei, L. ImageNet: A large-scale hierarchical image database. *En 2009 IEEE Conference on Computer Vision and Pattern Recognition*, págs. 248–255. 2009. doi:10.1109/CVPR.2009.5206848.

[Deng14] Deng, L., Yu, D., et al. Deep learning: methods and applications. *Foundations and trends® in signal processing*, 7(3–4):197–387, 2014.

[Devlin22] Devlin, J., Chang, M.-W., Lee, K., y Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *En Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, págs. 4171–4186. 2022.

[Ding21] Ding, X., Hao, T., Liu, J., Han, J., Guo, Y., y Ding, G. Towards efficient model compression via learned global ranking. *En Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, págs. 1518–1528. 2021.

[Donahue14a] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., y Darrell, T. Decaf: A deep convolutional activation feature for generic visual recognition. *En International conference on machine learning*, págs. 647–655. PMLR, 2014.

[Donahue14b] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., y Darrell, T. Decaf: A deep convolutional activation feature for generic visual recognition. *International conference on machine learning*, págs. 647–655, 2014.

[Dong22] Dong, Y., Xie, K., Tong, X., y Xu, K. Towards robust neural networks via sparsification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5811–5826, 2022.

[Dosovitskiy16] Dosovitskiy, A. y Brox, T. Generating images with perceptual similarity metrics based on deep networks. *Advances in neural information processing systems*, 29, 2016.

[Douze16] Douze, M., Jégou, H., y Perronnin, F. Polysemous codes. *En Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, págs. 785–801. Springer, 2016.

[Douze24] Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.-E., Lomeli, M., Hosseini, L., y Jégou, H. The faiss library. *NA*, 2024.

[Du20] Du, X., Lin, T.-Y., Jin, P., Ghiasi, G., Tan, M., Cui, Y., Le, Q. V., y Song, X. Spinenet: Learning scale-permuted backbone for recognition and localization. *En Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, págs. 11592–11601. 2020.

[Fu19] Fu, C., Xiang, C., Wang, C., y Cai, D. Fast approximate nearest neighbor search with the navigating spreading-out graph. *Proceedings of the VLDB Endowment*, 12(5):461–474, 2019.

[Galushkin07] Galushkin, A. I. *Neural networks theory*. Springer Science & Business Media, 2007.

[Ge13] Ge, T., He, K., Ke, Q., y Sun, J. Optimized product quantization. *En Proceedings of the IEEE conference on computer vision and pattern recognition*, págs. 2946–2953. 2013.

[Gonzalez02] Gonzalez, R. C. y Woods, R. E. *Digital image processing*. Prentice hall Upper Saddle River, NJ, 2002.

[Goodfellow14] Goodfellow, I. J., Shlens, J., y Szegedy, C. Explaining and harnessing adversarial examples. *En International Conference on Learning Representations*. 2014.

[Gordo17] Gordo, A., Almazan, J., Revaud, J., y Larlus, D. End-to-end learning of deep visual representations for image retrieval. *International Journal of Computer Vision*, 124(2):237–254, 2017.

[Guo16] Guo, C. y Berkhahn, F. Entity embeddings of categorical variables. *arXiv preprint arXiv:1604.06737*, 2016.

[Guo20a] Guo, M., Yang, Z., Xu, D., Shen, Y., y Yu, H. Cp-nas: Child-parent neural architecture search for 1-bit cnns. *En International Joint Conference on Artificial Intelligence*, págs. 1452–1458. 2020.

[Guo20b] Guo, R., Sun, P., Lindgren, E., Geng, Q., Simcha, D., Chern, F., y Kumar, S. Accelerating large-scale inference with anisotropic vector quantization. *En International Conference on Machine Learning*, págs. 3887–3896. PMLR, 2020.

[Guyon03] Guyon, I. y Elisseeff, A. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[Hamming50] Hamming, R. W. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160, 1950. doi:10.1002/j.1538-7305.1950.tb00463.x.

[Hanzo11] Hanzo, L., Liew, T. H., Yeap, B. L., Tee, R. Y. S., y Ng, S. X. *Turbo Coding, Turbo Equalisation and Space-Time Coding: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels*. Wiley-

IEEE Press, 2ª ed$^{\text{ón}}$., 2011. ISBN 978-0-470-74726-9. doi:10.1002/9781119976554.

[He16a] He, K., Zhang, X., Ren, S., y Sun, J. Deep Residual Learning for Image Recognition. *En 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, págs. 770–778. 2016. doi:10.1109/CVPR.2016.90.

[He16b] He, K., Zhang, X., Ren, S., y Sun, J. Deep residual learning for image recognition. *En Proceedings of the IEEE conference on computer vision and pattern recognition*, págs. 770–778. 2016.

[He23] He, K., Zhang, X., Ren, S., y Sun, J. Deep residual learning for image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):3349–3364, 2023.

[Hedayat78] Hedayat, A. y Wallis, W. D. Hadamard matrices and their applications. *The annals of statistics*, págs. 1184–1238, 1978.

[Hernández17] Hernández, E. A., Alonso, M. A., Chávez, E., Covarrubias, D. H., y Conte, R. Robust polygon recognition method with similarity invariants applied to star identification. *Advances in Space Research*, 59(4):1095–1111, 2017.

[Horadam07] Horadam, K. J. *Hadamard matrices and their applications*. Princeton university press, Princeton, NJ, 2007. ISBN 9780691119212.

[Horadam12] Horadam, K. J. *Hadamard matrices and their applications*. Princeton university press, 2012.

[Hoyos21] Hoyos, A., Ruiz, U., y Chavez, E. Hadamard's defense against adversarial examples. *IEEE Access*, 9:118324–118333, 2021. doi:10.1109/ACCESS.2021.3106855.

[Iwasaki18] Iwasaki, M. y Miyazaki, D. Optimization of indexing based on k-nearest neighbor graph for proximity search in high-dimensional data. *arXiv preprint arXiv:1810.07355*, 2018.

[Jain96] Jain, A. K. y Vailaya, A. Image retrieval using color and shape. *Pattern recognition*, 29(8):1233–1244, 1996.

[Jain10] Jain, A. K., Duin, R. P., y Mao, J. *Introduction to pattern recognition and machine learning*. World Scientific Publishing Company, 2010.

[Janick19] Janick, J. The apple in history. *Horticultural Reviews*, 47:91–137, 2019.

[Jiang24] Jiang, Z. y Zhou, C. Comprehensive study on shape representation methods for shape-based object recognition. *Journal of Optics*, 53(3):1890–1896, 2024.

[Johnson19] Johnson, J., Douze, M., y J'egou, H. Billion-scale similarity search with gpus. *En IEEE Transactions on Big Data*. IEEE, 2019.

[Johnson23] Johnson, W. B., Lindenstrauss, J., y Indyk, P. Fast similarity search with locality-sensitive hashing for deep learning models. *Journal of Machine Learning Research*, 24(103):1–32, 2023.

[Jégou11] Jégou, H., Douze, M., y Schmid, C. Product quantization for nearest neighbor search. *En IEEE transactions on pattern analysis and machine intelligence*, tomo 33, págs. 117–128. IEEE, 2011.

[Keogh09] Keogh, E., Wei, L., Xi, X., Vlachos, M., Lee, S.-H., y Protopapas, P. Supporting exact indexing of arbitrarily rotated shapes and periodic time series under euclidean and warping distance measures. *The VLDB journal*, 18(3):611–630, 2009.

[Kim00] Kim, W.-Y. y Kim, Y.-S. A region-based shape descriptor using zernike moments. *Signal processing: Image communication*, 16(1-2):95–102, 2000.

[Kirtas23] Kirtas, M., Passalis, N., Oikonomou, A., Moralis-Pegios, M., Giamougiannis, G., Tsakyridis, A., Mourgias-Alexandris, G., Pleros, N., y Tefas, A. Mixed-precision quantization-aware training for photonic neural networks. *Neural Computing and Applications*, págs. 1–19, 2023.

[Kloberdanz22] Kloberdanz, E. y Le, W. Mixquant: A quantization bit-width search that can optimize the performance of your quantization method. *not*, 2022.

[Kratochvíl20] Kratochvíl, M., Veselý, P., Mejzlík, F., y Lokoč, J. SOM-Hunter: Video Browsing with Relevance-to-SOM Feedback Loop. *En* Y. M. Ro, W.-H. Cheng, J. Kim, W.-T. Chu, P. Cui, J.-W. Choi, M.-C. Hu, y W. De Neve, eds., *MultiMedia Modeling*, págs. 790–795. Springer International Publishing, Cham, 2020. ISBN 978-3-030-37734-2.

[Krizhevsky09] Krizhevsky, A. y Hinton, G. Learning multiple layers of features from tiny images. Inf. Téc. 0, University of Toronto, Toronto, Ontario, 2009.

[Krizhevsky12a] Krizhevsky, A., Sutskever, I., y Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *En* F. Pereira, C. J. C. Burges, L. Bottou, y K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems*, tomo 25. Curran Associates, Inc., 2012.

[Krizhevsky12b] Krizhevsky, A., Sutskever, I., y Hinton, G. E. Imagenet classification with deep convolutional neural networks. *En Advances in neural information processing systems*, págs. 1097–1105. 2012.

[Kumar21] Kumar, R. y Mali, K. Shape classification via contour matching using the perpendicular distance functions. *International Journal of Engineering and Applied Physics*, 1(2):192–198, 2021.

[Latecki00] Latecki, L. J., Lakamper, R., y Eckhardt, T. Shape descriptors for non-rigid shapes with a single closed contour. *En Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, tomo 1, págs. 424–429. IEEE, 2000.

[Latif19] Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., Zafar, B., Dar, S. H., Sajid, M., y Khalil, T. Content-based image retrieval and feature extraction: A comprehensive review. *Mathematical problems in engineering*, 2019(1):9658350, 2019.

[LeCun10] LeCun, Y., Kavukcuoglu, K., y Farabet, C. Convolutional networks and applications in vision. *En Proceedings of 2010 IEEE international symposium on circuits and systems*, págs. 253–256. IEEE, 2010.

[LeCun15] LeCun, Y., Bengio, Y., y Hinton, G. *Deep learning.* nature, 2015.

[Li20] Li, K. y Li, G.-N. Approximate nearest neighbor search: A comprehensive survey. *Proceedings of the VLDB Endowment*, 13(11):2852–2886, 2020.

[Li21] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., y Liu, H. Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6):1–45, 2021.

[Li22a] Li, H., Kadav, A., Durdanovic, I., y Samet, H. Neuron selection and pruning for efficient deep neural networks. *En Proceedings of the 39th International Conference on Machine Learning*, págs. 5821–5830. 2022.

[Li22b] Li, Z., Gong, B., y Yang, T. Knowledge transfer in deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5872–5891, 2022.

[Lin14] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., y Zitnick, C. L. Microsoft coco: Common objects in context. *En* D. Fleet, T. Pajdla, B. Schiele, y T. Tuytelaars, eds., *Computer Vision – ECCV 2014*, págs. 740–755. Springer International Publishing, Cham, 2014. ISBN 978-3-319-10602-1.

[Lin17] Lin, T.-Y., Goyal, P., Girshick, R., He, K., y Dollár, P. Focal loss for dense object detection. *En Proceedings of the IEEE international conference on computer vision*, págs. 2980–2988. 2017.

[Liu07] Liu, Y., Zhang, D., Lu, G., y Ma, W.-Y. A survey of content-based image retrieval with high-level semantics. *Pattern recognition*, 40(1):262–282, 2007.

[Liu22a] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., et al. Swin transformer v2: Scaling up capacity and resolution. *En Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, págs. 12009–12019. 2022.

[Liu22b] Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., y Xie, S. A convnet for the 2020s. *CoRR*, abs/2201.03545, 2022.
URL `https://arxiv.org/abs/2201.03545`

[Liu23] Liu, S., Mocanu, D. C., Pei, Y., y Pechenizkiy, M. Dynamic sparse training: Find efficient sparse network from scratch with trainable masked layers. *En International Conference on Learning Representations*. 2023.

[Lowe04] Lowe, D. G. Distinctive image features from scale-invariant key-

points. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[MacWilliams77] MacWilliams, F. J. y Sloane, N. J. A. *The theory of error-correcting codes*, tomo 16. Elsevier, 1977.

[Malkov18] Malkov, Y. A. y Yashunin, D. A. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):824–836, 2018.

[Malkov20a] Malkov, Y. A. y Yashunin, D. A. Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):824–836, 2020. doi:10.1109/TPAMI.2018.2889473.

[Malkov20b] Malkov, Y. A. y Yashunin, D. A. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):824–836, 2020.

[Manjunath96] Manjunath, B. S. y Ma, W.-Y. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842, 1996.

[Manning08] Manning, C. D., Raghavan, P., y Schütze, H. *Introduction to information retrieval*. Cambridge University Press, Cambridge, 2008.

[Menghani23] Menghani, G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys*, 55(12):1–37, 2023.

[Mikolajczyk05] Mikolajczyk, K. y Schmid, C. A performance evaluation of local

descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[Mikolov13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., y Dean, J. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26, 2013.

[Mikolov23] Mikolov, T., Sutskever, I., y Chen, K. Advances in embedding techniques for natural language processing. *Journal of Artificial Intelligence Research*, 68:1–43, 2023.

[Mingqiang08] Mingqiang, Y., Kidiyo, K., Joseph, R., et al. A survey of shape feature extraction techniques. *Pattern recognition*, 15(7):43–90, 2008.

[Mishra21] Mishra, R. K., Reddy, G. S., y Pathak, H. The understanding of deep learning: A comprehensive review. *Mathematical Problems in Engineering*, 2021(1):5548884, 2021.

[Mohammed23] Mohammed, M. A., Hussain, M. A., Oraibi, Z. A., Abduljabbar, Z. A., y Nyangaresi, V. O. Secure content based image retrieval system using deep learning. *J. Basrah Res.(Sci.)*, 49(2):94–111, 2023.

[Mokhtarian97] Mokhtarian, F., Abbasi, S., y Kittler, J. Efficient and robust retrieval by shape content through curvature scale space. *En Image Databases and Multi-Media Search*, págs. 51–58. World Scientific, 1997.

[Moon20] Moon, T. K. *Error correction coding: mathematical methods and algorithms*. John Wiley & Sons, 2020.

[Muja14] Muja, M. y Lowe, D. G. Scalable nearest neighbor algorithms for

high dimensional data. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2227–2240, 2014.

[Pan09] Pan, S. J. y Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[Pan10] Pan, S. J. y Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. doi:10.1109/TKDE.2009.191.

[Pan24] Pan, J. J., Wang, J., y Li, G. Vector database management techniques and systems. *En Companion of the 2024 International Conference on Management of Data*, págs. 597–604. 2024.

[Paramarthalingam21a] Paramarthalingam, A. y Thankanadar, M. Extraction of compact boundary normalisation based geometric descriptors for affine invariant shape retrieval. *IET Image Processing*, 15(5):1093–1104, 2021.

[Paramarthalingam21b] Paramarthalingam, A. y Thankanadar, M. Extraction of compact boundary normalisation based geometric descriptors for affine invariant shape retrieval. *IET Image Processing*, 15(5):1093–1104, 2021.

[Park09] Park, S. Y. y Bera, A. K. The minimum cross-entropy principle for probability distribution function estimation. *Journal of Econometrics*, 150(2):219–230, 2009.

[Parola21] Parola, M., Nannini, A., y Poleggi, S. Web image search engine based on lsh index and cnn resnet50. *arXiv preprint arXiv:2108.13301*, 2021.

[Proakis08] Proakis, J. G. y Salehi, M. *Digital communications*. McGraw-hill, 2008.

[Quiroz24] Quiroz, B., Martinez, B., Camarena-Ibarrola, A., y Chavez, E. Design of a brief perceptual loss function with hadamard codes. *Multimedia Tools and Applications*, págs. 1–20, 2024.

[Radford21] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning transferable visual models from natural language supervision. *En International conference on machine learning*, págs. 8748–8763. PMLR, 2021.

[Radosavovic20] Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., y Dollár, P. Designing network design spaces, 2020.

[Redmon16] Redmon, J., Divvala, S., Girshick, R., y Farhadi, A. You only look once: Unified, real-time object detection. *En Proceedings of the IEEE conference on computer vision and pattern recognition*, págs. 779–788. 2016.

[Reed60] Reed, I. S. y Solomon, G. Polynomial codes over certain finite fields. *Journal of The Society for Industrial and Applied Mathematics*, 8:300–304, 1960.

[Robertson09] Robertson, S., Zaragoza, H., et al. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4):333–389, 2009.

[Rui99] Rui, Y., Huang, T. S., y Chang, S.-F. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39–62, 1999.

[Sablayrolles19] Sablayrolles, A., Douze, M., Schmid, C., y J'egou, H. Spreading vectors for similarity search. *En International Conference on Learning Representations*. 2019.

[Salton83] Salton, G. y McGill, M. J. *Introduction to modern information retrieval*. McGraw-Hill, New York, 1983.

[Schuhmann22] Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C. W., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S. R., Crowson, K., Schmidt, L., Kaczmarczyk, R., y Jitsev, J. LAION-5b: An open large-scale dataset for training next generation image-text models. *En Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. 2022.
URL `https://openreview.net/forum?id=M3Y74vmsMcY`

[Shannon48] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

[Simonyan15] Simonyan, K. y Zisserman, A. Very deep convolutional networks for large-scale image recognition. *En* Y. Bengio y Y. LeCun, eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. 2015.

[Smeulders00] Smeulders, A. W., Worring, M., Santini, S., Gupta, A., y Jain, R. Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1349–1380, 2000.

[Sparck Jones72] Sparck Jones, K. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972.

[Subramanya20] Subramanya, S. J., Devvrit, R., Simhadri, H. V., Krishnawamy, R., y Kadekodi, A. Diskann: Fast accurate billion-point nearest

neighbor search on a single node. *En Proceedings of the 33rd International Conference on Neural Information Processing Systems*, págs. 13748–13758. 2020.

[Sun20] Sun, Y., Xie, Y., Guo, P., Yu, J., Zheng, R., Zhu, C., Lin, Y., Wu, B., Jiang, X., y Wang, Y. Accelerating cnn training by pruning activation gradients. *En European Conference on Computer Vision*, págs. 322–338. Springer, 2020.

[Swain91] Swain, M. J. y Ballard, D. H. Color indexing. *International journal of computer vision*, 7(1):11–32, 1991.

[Sylvester67] Sylvester, J. J. Thoughts on inverse orthogonal matrices, simultaneous sign-successions, and tessellated pavements in two or more colours, with applications to newton's rule, ornamental tile-work, and the theory of numbers. *Philosophical Magazine and Journal of Science*, 34(232):461–475, 1867.

[Talamantes22] Talamantes, A. y Chavez, E. Instance-based learning using the half-space proximal graph. *Pattern Recognition Letters*, 156:88–95, 2022.

[Tan19a] Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., y Le, Q. V. Mnasnet: Platform-aware neural architecture search for mobile, 2019.

[Tan19b] Tan, M. y Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *En International Conference on Machine Learning*, págs. 6105–6114. PMLR, 2019.

[Tu22] Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., y Li, Y. Maxvit: Multi-axis vision transformer. *ECCV*, 2022.

[Vadera22] Vadera, S. y Ameen, S. Methods for pruning deep neural networks. *IEEE Access*, 10:63280–63300, 2022.

[Vaswani23] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., y Polosukhin, I. Attention is all you need. *Journal of Machine Learning Research*, 24(138):1–38, 2023.

[Vinyals16] Vinyals, O., Blundell, C., Lillicrap, T., kavukcuoglu, k., y Wierstra, D. Matching networks for one shot learning. *En* D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, y R. Garnett, eds., *Advances in Neural Information Processing Systems*, tomo 29. Curran Associates, Inc., 2016.

[Wan14] Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., y Li, J. Deep learning for content-based image retrieval: A comprehensive study. *En Proceedings of the 22nd ACM international conference on Multimedia*, págs. 157–166. 2014.

[Wang21] Wang, J., Zhang, T., Song, J., Sebe, N., y Shen, H. T. A comprehensive survey on fast similarity search techniques. *ACM Computing Surveys (CSUR)*, 54(5):1–37, 2021.

[Wang22a] Wang, C., Zhang, G., y Grosse, R. Neuron-level structured pruning using polarization regularizer. *En Advances in Neural Information Processing Systems*, tomo 35, págs. 15101–15114. 2022.

[Wang22b] Wang, J., Chen, L., y Zhang, W. Efficient nearest neighbor search in high-dimensional spaces: Techniques and future directions. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3456–3470, 2022.

[Wang22c] Wang, Y., Zhang, H., Li, X., y Liu, X. Applications of nearest neighbor search: A survey. *ACM Computing Surveys*, 55(3):1–36, 2022.

[Wang23a] Wang, J., Zhang, T., y Liu, J. A comprehensive survey of dimensionality reduction techniques in deep learning. *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, 45(6):6712–6738, 2023.

[Wang23b] Wang, Z., Duan, T., Fang, L., Suo, Q., y Gao, M. Efficient visual representation learning with contrastive masked autoencoders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10879–10892, 2023.

[Weber98] Weber, R., Schek, H.-J., y Blott, S. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. *VLDB*, 98:194–205, 1998.

[Weiss16] Weiss, K., Khoshgoftaar, T. M., y Wang, D. A survey of transfer learning. *Journal of Big data*, 3:1–40, 2016.

[Xu08] Xu, X., Lee, D.-J., Antani, S., y Long, L. R. A spine x-ray image retrieval system using partial shape matching. *IEEE Transactions on Information Technology in Biomedicine*, 12(1):100–108, 2008.

[Xu22] Xu, Y., Liu, J., y Chen, Q. A survey on efficient nearest neighbor search in high-dimensional spaces. *ACM Computing Surveys*, 55(3):1–35, 2022.

[Yeom21] Yeom, S.-W., Seegerer, P., Lapuschkin, S., Binder, A., Wiedemann, S., Müller, K.-R., y Samek, W. Pruning by explaining: A novel criterion for deep neural network pruning. *En Pattern Recognition*, págs. 119–133. Springer, 2021.

[Yildirim21] Yildirim, M. E., Ince, O. F., Yucel, B. S., y Ince, I. F. Shape retrieval using angle-wise contour variance. *Journal of Electrical Engineering*, 72(2):99–105, 2021.

[Yosinski14] Yosinski, J., Clune, J., Bengio, Y., y Lipson, H. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27, 2014.

[Yu21] Yu, S., Giraldo, L. G. S., y Príncipe, J. C. Information-theoretic methods in deep neural networks: Recent advances and emerging opportunities. *En IJCAI*, págs. 4669–4678. 2021.

[Zebari20] Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., y Saeed, J. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(1):56–70, 2020.

[Zeiler14] Zeiler, M. D. y Fergus, R. Visualizing and understanding convolutional networks. *En European conference on computer vision*, págs. 818–833. Springer, 2014.

[Zeng22] Zeng, A., Attarian, M., Ichter, B., Choromanski, K., Wong, A., Welker, S., Tombari, F., Purohit, A., Ryoo, M., Sindhwani, V., et al. Socratic models: Composing zero-shot multimodal reasoning with language. *arXiv preprint arXiv:2204.00598*, 2022.

[Zhang03] Zhang, D. y Lu, G. Evaluation of mpeg-7 shape descriptors against other shape descriptors. *multimedia systems*, 9(1):15–30, 2003.

[Zhang04] Zhang, D. y Lu, G. Review of shape representation and description techniques. *Pattern recognition*, 37(1):1–19, 2004.

[Zhang21] Zhang, C., Zheng, Y., Guo, B., Li, C., y Liao, N. Scn: a novel shape classification algorithm based on convolutional neural network. *Symmetry*, 13(3):499, 2021.

[Zhang22] Zhang, Y. y LeCun, Y. Efficient dimensionality reduction for image classification using deep learning. *En Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, págs. 2345–2354. 2022.

[Zhao24] Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., y Parmar,

M. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57(4):99, 2024.

[Zhong16] Zhong, G., Wang, L.-N., Ling, X., y Dong, J. An overview on data representation learning: From traditional feature learning to recent deep learning. *The Journal of Finance and Data Science*, 2(4):265–278, 2016.

Bryan Eduardo Martínez Guzmán

DISEÑO DE ÍNDICES PARA LA RECUPERACIÓN DE OBJETOS MULTIMEDIA.

Universidad Michoacana de San Nicolás de Hidalgo

Detalles del documento

Identificador de la entrega
trn:oid:::3117:419653321

Fecha de entrega
7 ene 2025, 1:21 p.m. GMT-6

Fecha de descarga
7 ene 2025, 1:32 p.m. GMT-6

Nombre de archivo
DISEÑO DE ÍNDICES PARA LA RECUPERACIÓN DE OBJETOS MULTIMEDIA.pdf

Tamaño de archivo
15.2 MB

195 Páginas

48,225 Palabras

269,886 Caracteres

15% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Fuentes principales

14% Fuentes de Internet
10% Publicaciones
0% Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alerta de integridad para revisión

Caracteres reemplazados
85 caracteres sospechosos en N.º de páginas
Las letras son intercambiadas por caracteres similares de otro alfabeto.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Formato de declaración de originalidad y uso de Inteligencia Artificial

Coordinación General de Estudios de Posgrado
Universidad Michoacana de San Nicolás de Hidalgo

Morelia, Mich., a 6 de enero de 2025

A quien corresponda,

Por este medio, el/la abajo firmante, bajo protesta de decir verdad, declara lo siguiente:

- que presenta para revisión el manuscrito cuyos detalles se especifican abajo.
- que todas las fuentes consultadas para la elaboración del manuscrito están debidamente identificadas dentro del cuerpo del texto, e incluidas en la lista de referencias.
- que, en caso de haber usado un sistema de inteligencia artificial, en cualquier etapa del desarrollo de su trabajo, lo ha especificado en la tabla que se encuentra en este documento.
- que conoce la normativa de la Universidad Michoacana de San Nicolás de Hidalgo, en particular los Incisos IX y XII del artículo 85, y los artículos 88 y 101 del Estatuto Universitario de la UMSNH, además del transitorio tercero del Reglamento General para los Estudios de Posgrado de la UMSNH.

Atentamente,

C. _____

Datos del manuscrito que se presenta a revisión		
Programa educativo	Doctorado en Ciencias en Ingeniería Eléctrica opción Sistemas Computacionales	
Título del trabajo	Diseño de índices para la recuperación de objetos multimedia	
	Nombre	**Correo electrónico**
Autor/es	Bryan Eduardo Martínez Guzmán	1007904a@umich.mx
Director	José Antonio Camarena Ibarrola	antonio.camarena@umich.mx
Codirector	Edgar Leonel Chávez González	elchavez@cicese.mx
Coordinador del programa	José Antonio Camarena Ibarrola	antonio.camarena@umich.mx

Uso de Inteligencia Artificial		
Rubro	**Uso (sí/no)**	**Descripción**
Asistencia en la redacción	Sí	En mis textos primero escribo varios párrafos e ideas, luego uso una IA para mejorarlos, y los vuelvo a revisar por mi cuenta
Traducción al español	Sí	Uso el traductor de google y uno que se basa en deep learning
Traducción a otra lengua	No	N/A
Revisión y corrección de estilo	No	Siempre trato de usar mi estilo, solo uso claude.ai para revisión de ideas y su entendimiento
Análisis de datos	No	Uso técnicas de análisis del estado del arte, así como propuestas por otros investigadores
Búsqueda y organización de información	Sí	Uso claude.ai para marcar el inicio de lo que deseo conocer, luego busco en google scholar los trabajos más recientes e importantes relacionados
Formateo de las referencias bibliográficas	No	Uso google scholar para las referencias en bibtex
Generación de contenido multimedia	No	Solo he usado bases de datos de dominio público como lo es Imagenet, YFCC100M, Coco ... etc
Otro		