

Construcción de una Función de Distancia para Consultar por Similitud Caracteres de Hueso de Oráculo

Andrés J. Pascal

*FRCU - Universidad Tecnológica Nacional
pascala@frcu.utn.edu.ar*

Adrián N. Planas

*FRCU - Universidad Tecnológica Nacional
planasa@frcu.utn.edu.ar*

Federico Stauber

*FRCU - Universidad Tecnológica Nacional
fedestauber@gmail.com*

Martín López

*FRCU - Universidad Tecnológica Nacional
tincho1597lopez@gmail.com*

León Castiglioni

*FRCU - Universidad Tecnológica Nacional
leoncastiglioni4@gmail.com*

Resumen

Los caracteres de hueso de oráculo (OBC) de la antigua China representan el sistema de escritura antiguo más renombrado a nivel mundial. El estudio e identificación de los OBC y su desciframiento se erigen como uno de los aspectos más cruciales dentro de la esfera de investigación de estos artefactos históricos. Entre los desafíos que enfrenta esta investigación, destaca el hecho de que la revisión de la literatura al respecto suele demandar considerables recursos temporales y de mano de obra. En consecuencia, la digitalización de la literatura OBC surge como una dirección inevitable para el desarrollo futuro de este campo. Por otro lado, durante la última década las Redes Neuronales Profundas Convolucionales (CNN) han demostrado un alto rendimiento en el procesamiento automático de imágenes. Este artículo presenta un enfoque que combina preprocesamiento, aumento de datos y modelos CNN para aprender una función de distancia para buscar por similitud caracteres OBC en un escenario Few-Shot Learning, utilizando una arquitectura de Redes Siamesas en su proceso de entrenamiento. La principal ventaja de utilizar Búsquedas por Similitud en lugar de modelos de clasificación, es que el sistema permite el agregado de nuevos elementos (clases) sin modificación del modelo ni reentrenamiento.

Palabras Clave: Búsquedas por Similitud, OBC, Metric Learning, CNNs, Redes Siamesas, Few-Shot Learning, Aumentación.

1. Introducción

La expansión de tipos de datos que abarcan imágenes, audio, video y texto ha evidenciado la

limitación de los modelos de búsqueda convencionales. Como respuesta, la búsqueda por similitud emerge como una solución eficaz para localizar objetos análogos a una consulta en bases de datos no convencionales. Los Espacios Métricos [1] representan un enfoque formalizado que facilita búsquedas por similitud y la implementación de métodos de acceso más eficaces.

Este artículo se centra en el desarrollo de técnicas de aprendizaje de funciones de distancia con el fin de facilitar la búsqueda de similitud en imágenes que representan caracteres óseos de oráculo (OBC). Estos caracteres se remontan a la Edad del Bronce en China y representan el sistema de escritura más antiguo documentado en el país [2]. Una de las principales dificultades abordadas en esta investigación se relaciona con la identificación de los glifos dentro de los OBC. Desde la perspectiva de la temática estudiada, se reconoce que los glifos de los OBC son intrincados y diversos, y un solo OBC puede presentar múltiples variantes, lo que resulta en un significativo desafío en términos de inversión de tiempo y recursos humanos. Desde la perspectiva metodológica, la investigación de los glifos de OBC implica un exhaustivo análisis de la literatura existente; sin embargo, los libros impresos sobre este tema suelen ser densos en contenido y difíciles de recuperar, lo que agrega complejidad al proceso investigativo.

Recientemente, el campo de la visión por computadora ha experimentado un avance sustancial gracias a las técnicas de Redes Neuronales Convolucionales Profundas (DCNN). Los notables logros alcanzados por las DCNN en la comunidad de visión por computadora motivan a considerar su aplicación en la búsqueda de Caracteres Óseos de Oráculo (OBC). Sin embargo, estos enfoques basados en DCNN dependen en gran medida de conjuntos de

datos de entrenamiento con etiquetas, recursos que a menudo resultan inaccesibles, particularmente en este contexto específico.

Este problema plantea tres desafíos sustanciales en el ámbito del aprendizaje automático: a) se dispone de pocas muestras (unas 20) por categoría (Few-Shot Learning), b) a medida en que se descubren nuevos OBC, la cantidad de categorías aumenta, y c) la definición de una función de distancia precisa para medir la similitud entre dos imágenes es una tarea compleja. En este artículo, se exponen modelos de Redes Neuronales Convolucionales (CNN) junto con técnicas de preprocesamiento, aumento de datos y entrenamiento, con el propósito de crear una función de distancia efectiva que evalúe el nivel de similitud entre imágenes de caracteres óseos de oráculo (OBC).

El resto de este documento está organizado de la siguiente manera: la Sección 2 presenta el trabajo relacionado; en la Sección 3, se explica el proceso de aprendizaje de la función de distancia y los modelos utilizados. La Sección 4 describe los experimentos realizados y la 5 los resultados obtenidos. Finalmente, en la Sección 6 se presentan las conclusiones y el trabajo futuro.

2. Trabajo Relacionado

En esta sección, se expone el contexto del estudio actual, destacando las técnicas fundamentales empleadas para emplear la Redes Siamesas basadas en CNNs como medida de similitud de comparación de OBCs.

2.1 Recuperación de Imágenes Basada en Contenido

La Recuperación de Imágenes Basada en Contenido (CBIR) es un procedimiento que busca recuperar imágenes de una base de datos considerando características visuales como similitud en colores o formas. Los atributos predominantes son color, textura y forma [3, 4]. El color prevalece debido a su facilidad para ser extraído [5]. En contraste, adquirir datos sobre forma y textura es complejo y costoso [6]. Este enfoque es crucial para aplicaciones que requieren la búsqueda eficiente de imágenes relevantes, como la medicina y la gestión de activos visuales. Mejorar la extracción de atributos complejos es un desafío continuo en el CBIR.

Los histogramas [7] son soluciones ampliamente empleadas, caracterizadas por su eficiencia computacional. Sin embargo, presentan insensibilidad a la variación de la posición de la cámara, carencia de información espacial y limitaciones en la adaptación a cambios en la luminosidad [8].

La forma es una característica visual fundamental utilizada para describir el contenido de las imágenes, pero su representación y descripción plantean desafíos significativos. La forma puede verse afectada por defectos, ruido, oclusión y distorsiones arbitrarias, lo

que dificulta su análisis. Además de esto, no se sabe qué características son más importantes en la forma. En resumen, la recuperación de imágenes basada en formas busca medir similitud entre formas representadas por sus características distintivas. Una forma puede ser descrita por diferentes aspectos [9] tales como: centro de gravedad (centroide) [10], masa, media, dispersión, varianza, eje de menor inercia, rectangularidad y convexidad. Para una mejor representación de formas, se han adoptado enfoques más avanzados, como el uso de descriptores invariantes, como Momentos de Hu, Legendre o Zernike [11, 12, 13], que ofrecen resultados más confiables y precisos en el análisis de formas complejas.

2.2 Búsquedas en Espacios Métricos

Los Sistemas de Recuperación de Imágenes Basados en Contenido (CBIR) pueden ser generalizados y modelados mediante Espacios Métricos con el fin de lograr búsquedas eficientes. En el artículo [1], se demuestra que el problema de búsqueda de similitud puede ser formulado de la siguiente manera: dado un conjunto U de objetos y una función de distancia d definida entre ellos para cuantificar su similitud, el objetivo es recuperar todos los elementos similares a un objeto dado utilizando d como criterio. Esta formulación permite desarrollar enfoques y algoritmos que facilitan la identificación y recuperación eficiente de imágenes similares en grandes bases de datos, lo que resulta fundamental en diversas aplicaciones prácticas y académicas.

Esta función d satisface las propiedades requeridas para ser una métrica:

- (a) $\forall x \in U, d(x, x) = 0$ (reflexividad)
- (b) $\forall x, y \in U, d(x, y) \geq 0$ (positividad)
- (c) $\forall x, y \in U, d(x, y) = d(y, x)$ (simetría)
- (d) $\forall x, y, z \in U, d(x, z) \leq d(x, y) + d(y, z)$ (desigualdad triangular)

Un subconjunto finito X de U , denominado base de datos, se utiliza para realizar la búsqueda. Uno de los tipos de consulta relevante en las búsquedas por similitud es la consulta de los k Vecinos Más Cercanos o $NN_k(q)_d$, que recupera los k elementos de X más cercanos a q :

$$NN_k(q)_d = A = \{x \in X / \forall y \in (X-A), d(q, x) \leq d(q, y)\} \text{ y } |A| = k$$

Dada una base de datos con n objetos, responder a estas consultas de forma trivial implica realizar n evaluaciones de distancia, lo cual puede resultar altamente costoso en aplicaciones prácticas. La relevancia de modelar estas consultas mediante espacios métricos se basa en la posibilidad de utilizar índices que aprovechan la propiedad de la desigualdad triangular para descartar elementos sin necesidad de compararlos

directamente con la consulta. Esto mejora significativamente el proceso de búsqueda [1, 14, 15], convirtiéndolo en una solución eficiente en diversas aplicaciones.

2.3 Redes Neuronales Convolucionales (DCNN o CNN) y Redes Siamesas

La arquitectura de las Redes Neuronales Convolucionales (CNNs) fue concebida por Fukushima en 1988 [16], mientras que en la década de los años '90, LeCun y otros investigadores [17] introdujeron algoritmos de aprendizaje basados en el gradiente que demostraron resultados alentadores en tareas de clasificación. Sin embargo, fue en la última década cuando las CNNs experimentaron avances notables y destacaron en una variedad de problemas de reconocimiento de patrones.

Ejemplos de arquitecturas actuales de CNN incluyen AlexNet [18], VGG Net [19], y NiN [20]. Además, se han propuesto alternativas avanzadas y eficientes como DenseNet [21], GoogLeNet [22], y Residual Networks [23]. Aunque los componentes básicos son similares en todas las arquitecturas, las variaciones topológicas conducen a diferencias significativas en la eficiencia del entrenamiento y la precisión de los resultados.

El algoritmo de aprendizaje de las Redes Neuronales Convolucionales (CNNs) no incorpora naturalmente el concepto de similitud entre imágenes. No obstante, arquitecturas recientes, como las Redes Neuronales Siamesas [24, 25] y Triplet Loss [26], han surgido para abordar este concepto y han demostrado eficacia en el reconocimiento facial. Las Redes Siamesas constan de dos CNNs idénticas, utilizadas para extraer vectores de características o entrenar funciones de distancia. Durante el entrenamiento, ajustan sus parámetros para minimizar la distancia entre imágenes similares y maximizarla entre imágenes diferentes, lo que facilita la representación y comparación de similitudes en el espacio de características.

La utilización de Redes Siamesas en conjunto con Redes Neuronales Convolucionales (CNNs) con el objetivo de extraer características destinadas a la Búsqueda por Similitud presenta dos desafíos críticos:

El primero radica en la limitada disponibilidad de instancias para el entrenamiento, un fenómeno conocido como One-Shot Learning o Few-Shots Learning [27, 28]. Esta limitación dificulta considerablemente el entrenamiento directo de los modelos. Hasta la fecha, la estrategia más efectiva para abordar esta dificultad ha sido el Transfer Learning y Fine Tuning [29, 30]. Sin embargo, esta técnica se ve restringida por la disponibilidad de bases de datos que contengan características similares.

El segundo desafío se relaciona con la generalización. Cuando una CNN se entrena para extraer vectores característicos, su comportamiento tiende a estar fuertemente ligado a las clases con las que

se entrenó inicialmente. Esto conlleva dificultades para adaptarse eficazmente a nuevas imágenes incorporadas a la base de datos sin un proceso de reentrenamiento. Una solución recientemente investigada es el Aprendizaje Métrico (Metric Learning) [31], que propone la obtención directa de una función de distancia correspondiente a los datos existentes en lugar de centrarse en la extracción de características. Éste es el enfoque que se emplea en el contexto del presente estudio.

2.4 Aprendizaje de Funciones de Distancia (Metric Learning)

El aprendizaje métrico [32] se centra en la adquisición de funciones de distancia que se ajustan a una definición de similitud específica. En el contexto del reconocimiento de patrones, esta definición de similitud es intrínseca a la tarea en cuestión, y el éxito del aprendizaje radica en la adecuada alineación con dicha tarea.

En [33], se realiza un análisis detallado de los métodos de aprendizaje métrico, proporcionando una visión panorámica de las diversas técnicas y enfoques empleados en este campo. En [34], se propone un enfoque de aprendizaje métrico basado en márgenes. En otro trabajo relacionado [35] se describe una técnica de aprendizaje métrico discriminatorio aplicada a la verificación de rostros. Por último, [36] ofrece una panorámica completa del aprendizaje métrico, abarcando múltiples enfoques y aplicaciones y subrayando los desafíos y tendencias prominentes en este dominio de investigación.

2.5 Búsquedas por Similitud de OBC

Investigaciones previas han abordado temas relacionados (clasificación) mediante diversos enfoques. En [37], se emplearon técnicas de visión por computadora para analizar caracteres de oráculo, generando el conjunto de datos Oracle-20K. Se estudiaron y analizaron representaciones visuales relacionadas con formas y bocetos, proponiendo una nueva representación jerárquica. Además, en [38] se integró el aprendizaje autosupervisado en el aumento de datos y se presentó un enfoque innovador, denominado Orc-Bert Augmentor, preentrenado para el reconocimiento de estos elementos en escenarios de datos escasos (Few-Shot).

En [39], se propone un método basado en una red neuronal convolucional profunda para el reconocimiento de OBC incompletos. Se extraen características del conjunto de datos de OBC incompletos con pocas muestras, permitiendo la identificación de OBC incompletos en diferentes imágenes. En otro trabajo [40], se utiliza una red neuronal convolucional (CNN) para mapear las imágenes de OBC a un espacio euclidiano donde la distancia entre las muestras mide sus similitudes,

posibilitando la clasificación mediante la regla del vecino más cercano (NN).

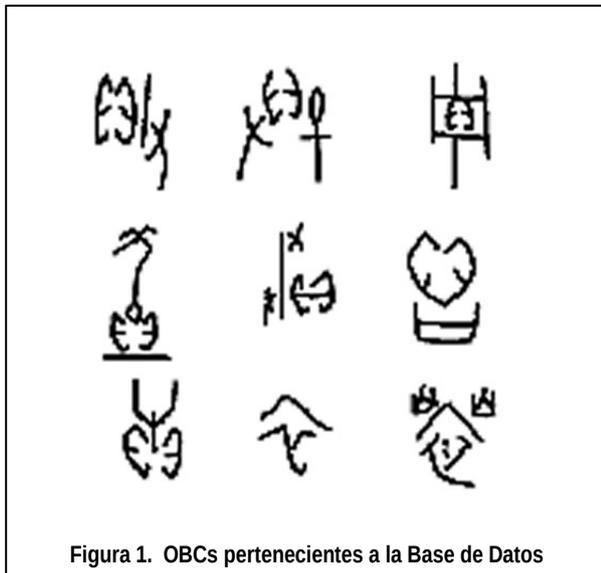
3. Aprendizaje de la Función de Distancia

En esta sección se describe la base de datos y los modelos utilizados para el aprendizaje de la función de distancia a partir de una base de datos compuesta por pocas muestras de cada OBC, utilizando Redes Siamesas, técnicas clásicas de Aumentación y el método de relieve presentado en [41].

3.1 Base de Datos de OBCs

En el presente estudio, se trabajó con una base de datos que consta de 1000 OBCs, con entre 19 y 24 muestras por cada uno. Cabe remarcar que estas imágenes fueron obtenidas de la base de datos pública llamada HWOBC, la cual está compuesta por caracteres chinos provenientes de la época de bronce, recopilados por investigadores y dibujados a mano por diferentes especialistas, para su uso de forma libre y gratuita.

En la Fig. 1 se exhiben nueve ejemplos de estos caracteres óseos de oráculo. Si bien algunos caracteres de huesos oraculares pueden haber tenido conexiones con objetos o conceptos específicos, no eran representaciones detalladas de esos objetos. Más bien, evolucionaron a partir de marcas y símbolos abstractos que se usaban para registrar eventos, fechas y preguntas en los huesos oraculares.



3.2 Modelos utilizados

En trabajos de investigación previos a esta problemática se probaron doce arquitecturas propias de CNNs a partir de las cuales se seleccionó el modelo de mejor rendimiento, que denominamos ModCar1 en este trabajo. ModCar1 es una CNN de seis capas

convolucionales, todas con activación ReLU. Cada capa convolucional está intercalada con una capa de Batch-Normalization. Este modelo utiliza un tamaño de vector de salida de 128 elementos. Las tres primeras capas convolucionales emplean 32 kernels, mientras que las restantes utilizan 64. El tamaño de los kernels es (3, 3) en cuatro capas y de (5, 5) en las dos restantes. Adicionalmente, se incorporan dos capas de Dropout para mitigar el overfitting.

Finalmente, el modelo incluye una capa "flatten", que representa la salida que es utilizada por la red siamesa. Este modelo posee 193.664 parámetros entrenables. Los vectores de salida se utilizan para obtener representaciones de las dos imágenes a comparar, que luego se concatenan y pasan por dos capas densas con función de activación ReLU. Por último, se agrega una capa extra con la función Sigmoide que devuelve el valor final de similitud entre ambas imágenes. Se emplea Binary Cross Entropy como función de pérdida y el optimizador Adam para el entrenamiento del modelo. Las imágenes utilizadas en la base de datos y en las consultas fueron de 50x50 píxeles.

Una característica importante de estas CNNs es que a pesar de ser profundas son relativamente pequeñas, lo que permite que su uso como función de distancia no sea excesivamente costoso.

A pesar de ello, hay que notar que este modelo no cumple todas las propiedades de una distancia métrica. De hecho, la función obtenida no es reflexiva, ni simétrica y tampoco cumple la desigualdad triangular. La simetría se resuelve fácilmente haciendo $d(x, y) = \text{CNN}(x, y) + \text{CNN}(y, x)$. Sin embargo, desde el punto de vista de los índices métricos, la propiedad que más importa es la desigualdad triangular, que es la que permite descartar elementos sin necesidad de compararlos con la consulta. Por esta razón, se realizó un experimento adicional para determinar en qué porcentaje la función de distancia satisface la desigualdad triangular para los elementos de la BD. Para ello se tomaron todas las ternas posibles y se verificó que más de un 90% de las mismas cumple esta propiedad para la función de distancia creada. Esto significa que sería posible utilizar índices métricos para acelerar la búsqueda, aun cuando en un porcentaje reducido de casos se podría estar descartando elementos erróneamente (índices métricos aproximados).

3.3 Preprocesamiento de Datos

Para salvar posibles diferencias entre los trazos de cada caracter, así también como posibles desplazamientos dentro del "lienzo" contenedor de la imagen, se decidió aplicar a las imágenes una serie de procesos de estandarización. Estas transformaciones fueron aplicadas en forma previa al ingreso de la imagen a la Red Siamesa. El primer paso fue la extracción de la subimagen correspondiente al "Minimun Bounding Box", para descartar parte del

fondo y con el objetivo de que el sistema sea robusto ante el escalado de las imágenes. El segundo proceso fue el reescalado de la misma a 50x50 con deformación, de tal manera de que los trazos ocupen todo el “lienzo” contenedor. Luego se eskeletonizaron y se aplicó la técnica de “Relieve” propuesta en [41], que consiste en aplicar una expansión linealmente decreciente con un alcance dado, hacia los lados de las líneas que componen la imagen. Esta técnica transforma la imagen de blanco y negro a escala de grises, y produce mejoras significativas en la eficacia del entrenamiento de imágenes cuando la cantidad de muestras es reducida. Por último, se reescalaron los valores de los pixels al intervalo [0, 1]. La Fig. 2 muestra dos caracteres sin procesar (primera fila) y sus correspondientes imágenes luego del procesamiento (segunda fila).

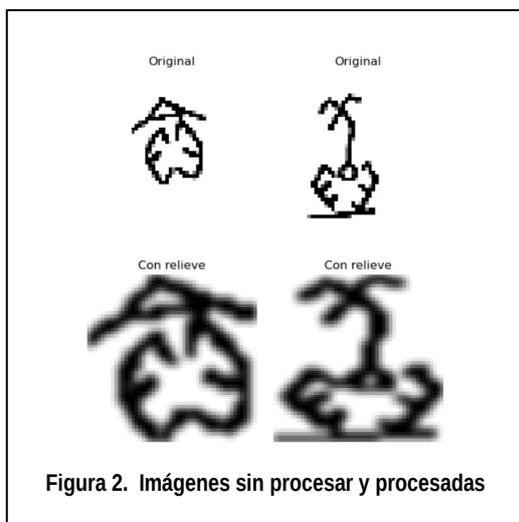


Figura 2. Imágenes sin procesar y procesadas

Los resultados experimentales demuestran que este mecanismo de preprocesamiento contribuye significativamente a que la Red Siamesa realice predicciones más robustas y estables.

4. Experimentos Realizados

Para el entrenamiento de la Red Siamesa se generaron al azar 200.000 pares de imágenes a partir de los elementos de la base de datos, utilizando preprocesamiento y aumentación estándar y asegurando un equilibrio entre la cantidad de pares similares y no similares. Para ello se seleccionaron 200 pares por cada imagen, 100 positivos y 100 negativos. El entrenamiento se realizó en 50 épocas alcanzando un 0,996 de precisión. Se utilizó una PC con procesador Ryzen 7 5800x, 32 GB RAM y GPU GeForce GTX 1080 con 2560 núcleos CUDA.

Como métrica principal para evaluar la función de distancia, se empleó la tasa de aciertos (Hit Rate, expresada en porcentaje), dado que es el indicador más representativo para determinar si la solución propuesta es aplicable en un contexto de búsquedas por similitud real.

Una vez entrenada la Red Siamesa como función de distancia, se realizaron búsquedas de los k vecinos más cercanos (NN_k) para probar su eficacia. Para ello distintos integrantes del equipo de investigación dibujaron un total de 50 imágenes de consulta, copiando con algunas variaciones caracteres existentes en la base de datos y luego las etiquetaron con el nombre de la respuesta esperada. Estos dibujos se llevaron a cabo íntegramente a “mano alzada” utilizando una aplicación de diseño gráfico.

Un segundo experimento se realizó para determinar el comportamiento de la función de distancia ante la incorporación de nuevas imágenes a la base de datos sin reentrenamiento. Para ello se agregaron 30 caracteres OBC que no formaban parte del conjunto inicial y se dibujaron otras 30 nuevas imágenes de consulta.

Los resultados se compararon con trabajos previos relacionados, aunque no sean exactamente de la misma naturaleza.

5. Resultados y Discusión

En esta sección se presentan y analizan los resultados de este estudio.

5.1 Eficacia de las Consultas por Similitud

Los porcentajes de acierto en las búsquedas por similitud de los k vecinos más cercanos (NN_k) utilizando el modelo propuesto ModCar1, para $k=1, 3$ y 5 , fueron: 90%, 98% y 100%. Estos valores se consideran suficientemente buenos para una aplicación real.

En la Tabla 1 se muestra la comparación de dichos resultados con los obtenidos en [41] mediante búsquedas por similitud también, pero sobre una base de datos distinta (marcas de ganado) y utilizando Redes Siamesas como mecanismo de extracción de características (modelo ConRelQD). Se eligió éste trabajo como punto de comparación dado que las imágenes sobre el cual se realizó son de la misma naturaleza y se aplicaron técnicas similares.

Tabla 1. Comparación de Porcentajes de Acierto

Modelo	NN_1	NN_3	NN_5
ConRelQD	-	82,50%	85,00%
ModCar1	90,00 %	98,00 %	100,00 %

El modelo ModCar1 supera significativamente (más de un 15% en todos los casos) el porcentaje de acierto del modelo presentado en el estudio de referencia. Sin embargo, hay que tener en cuenta que en éste caso el modelo es una función de distancia, mientras que ConRelQD es un modelo de extracción de

características que luego se utilizan como representaciones de las imágenes para ser comparadas a través de funciones de distancia métricas tales como la distancia euclidiana.

Una segunda comparación de resultados se realizó con el trabajo presentado en [2], que evalúa el rendimiento de cinco modelos convolucionales conocidos en la clasificación de caracteres OBC. Dichos modelos son AlexNet, ResNet, Vgg11, Cascada y Melnyk-Net. En la Tabla 2 se muestra la precisión promedio obtenida por cada uno de los modelos durante el entrenamiento y se comparan con ModCar1.

Tabla 2 – Comparación de la Precisión

Modelo	Precisión
AlexNet	96,62%
Vgg1 1	97,57%
ResNet	97,44%
Cascaded model	97,57%
Melnyk-Net	97,67%
ModCar1 (Similitud)	99,60 %

Nuevamente, hay que tener en cuenta que si bien en este caso la base de datos es la misma, ModCar1 es una Red Siamesa que determina si dos imágenes son similares o no lo son, mientras que las demás redes son modelos de clasificación o reconocimiento, que determinan a qué clase pertenece un elemento dado como entrada.

5.2 Eficacia de las Consultas Ante la Incorporación de Nuevos Elementos

Para evaluar el rendimiento de una función de distancia se debe estudiar el comportamiento ante la incorporación de nuevos objetos en la base de datos. Para ello se seleccionaron al azar 30 imágenes nuevas y se las agregaron al conjunto de datos inicial. Posteriormente se dibujaron a mano alzada 30 nuevas consultas, una por cada imagen, con las que se realizó el experimento. Cabe destacar que no se reentrenó la red, ya que el modelo debería funcionar para nuevos elementos (nuevas “clases”) sin ninguna modificación. Los resultados obtenidos se exponen en la Tabla 3.

Tabla 3 - Búsquedas de Nuevas Imágenes

NN ₁	NN ₃	NN ₅
70,00 %	90,00 %	100,00 %

A pesar de una disminución en los porcentajes observados en los casos de NN₁ y NN₃, es importante destacar que NN₃ continúa siendo significativamente elevado y que el rendimiento óptimo se mantiene invariable en el escenario de los 5 vecinos más cercanos.

6. Conclusión y Trabajo Futuro

Este artículo se centró en el desarrollo de técnicas de aprendizaje de funciones de distancia para facilitar la búsqueda de similitud en imágenes que representan caracteres óseos de oráculo (OBC). El estudio abordó los desafíos específicos asociados con la búsqueda de similitud en imágenes, que incluyen la disponibilidad limitada de muestras, el aumento constante de categorías a medida que se descubren nuevos OBC y la necesidad de definir una función de distancia precisa.

Se propuso un modelo de Red Siamesa que incluye una CNN, diseñado específicamente para abordar estos desafíos, y un mecanismo de preprocesamiento y aumento de datos que resultan eficaces en el entrenamiento de una función de distancia que determina el grado de similitud entre imágenes OBC.

Los resultados experimentales verificaron que el modelo ModCar1 logró tasas de acierto importantes en las búsquedas de los k vecinos más cercanos (NN _{k}), alcanzando valores entre el 90% y el 100% para distintos k . Estos resultados indican que el modelo es efectivo en la búsqueda de similitud en imágenes de caracteres OBC.

Estos resultados tienen aplicaciones potenciales en la identificación y búsqueda de glifos OBC en contextos de investigación y preservación del patrimonio.

Las tareas actuales y las actividades previstas para el futuro próximo son las siguientes:

- Realizar experimentos sobre la BD completa, que está compuesta por alrededor de 15.000 imágenes de OBCs.
- Explorar estrategias destinadas a mejorar la capacidad de generalización de la red, para que su rendimiento sea aún mayor frente a nuevas imágenes sin necesidad de un proceso de reentrenamiento.
- Evaluar el rendimiento de índices métricos aproximados para mejorar los tiempos de búsqueda.
- Aplicar el modelo propuesto con algunas variaciones para bases de datos de imágenes a color.

Referencias

- [1] Chávez, Edgar, et al. Searching in metric spaces. ACM computing surveys (CSUR) 33.3: 273-321 (2001).

- [2] LI, Bang, et al. HWOBC-a handwriting oracle bone character recognition database. En *Journal of Physics: Conference Series*. IOP Publishing, (2020). p. 012050.
- [3] Aslandogan, Y. Alp, and Clement T. Yu. Techniques and systems for image and video retrieval. *IEEE transactions on Knowledge and Data Engineering* 11.1: 56-63 (1999).
- [4] Smeulders, Arnold WM, et al. Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence* 22.12: 1349-1380 (2000).
- [5] Irena, V., Rachev, B., Vassilakopoulos, M. Optimization of the algorithm for image retrieval by color features. *CompSysTech*. (2006).
- [6] Sarfraz, M., Ahmad, R. Content-based image retrieval using multiple shape descriptors. *IEEE/ACS International Conference on Computer Systems and Applications* (2007).
- [7] Pass, G., Ramin Z. Histogram refinement for content-based image retrieval. *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96. IEEE*, (1996).
- [8] Zhang, HongJiang, et al. Image retrieval based on color features: An evaluation study. *Digital Image Storage and Archiving Systems*. Vol. 2606. SPIE, (1995).
- [9] Zhang, Dengsheng, and Guojun Lu. Review of shape representation and description techniques. *Pattern recognition* 37.1: 1-19 (2004).
- [10] Traina, Agma JM, et al. Content-based image retrieval using approximate shape of objects. *Proceedings. 17th IEEE Symposium on Computer-Based Medical Systems. IEEE*, (2004).
- [11] Celebi, M. Emre, and Y. Alp Aslandogan. A comparative study of three moment-based shape descriptors. *ITCC'05-Volume II*. Vol. 1. IEEE, (2005).
- [12] Zhang, Dengsheng, and Guojun Lu. Content-based shape retrieval using different shape descriptors: A comparative study. *ICME 2001. IEEE Computer Society*, (2001).
- [13] Li, Shan, Moon-Chuen Lee, and Chi-Man Pun. Complex Zernike moments features for shape-based image retrieval. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 39.1: 227-237 (2008).
- [14] Novak, David, Michal Batko, and Pavel Zezula. Metric index: An efficient and scalable solution for precise and approximate similarity search. *Information Systems* 36.4 (2011).
- [15] Britos, L., Printista, M., Reyes, N. DSACL+-tree: A dynamic data structure for similarity search in secondary memory. *ICSSA*. Springer, Berlin, Heidelberg (2012).
- [16] Fukushima, Kunihiko. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks* 1.2: 119-130 (1988).
- [17] LeCun, Yann, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86.11: 2278-2324 (1998).
- [18] Krizhevsky, Alex, Sutskever, Ilya and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the NIPS*. (2012).
- [19] Simonyan, Karen, and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [20] Lin, Min, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400* (2013).
- [21] Huang, Gao, et al. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017).
- [22] Szegedy, Christian, et al. Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. *Thirty-first AAAI conference on artificial intelligence*. (2017).
- [23] He, Kaiming, et al. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016).
- [24] Fierro, Atoany N., et al. Redes Convolucionales Siamesas y Tripletas para la Recuperación de Imágenes Similares en Contenido. *Información tecnológica* 30.6: 243-254 (2019).
- [25] Melekhov, Iaroslav, Juho Kannala, and Esa Rahtu. Siamese network features for image matching. *23rd international conference on pattern recognition (ICPR)*. IEEE, (2016).
- [26] Hoffer, Elad, and Nir Ailon. Deep metric learning using triplet network. *International workshop on similarity-based pattern recognition*. Springer, Cham, (2015).
- [27] Wang, Yaqing, et al. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)* 53.3: 1-34 (2020).
- [28] Lake, Brenden, et al. One shot learning of simple visual concepts. *Proceedings of the annual meeting of the cognitive science society*. Vol. 33. No. 33. (2011).
- [29] Pan, Sinno Jialin, and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22.10: 1345-1359 (2009).
- [30] Storkey, Amos. When training and test sets are different: characterizing learning transfer. *Dataset shift in machine learning* 30: 3-28 (2009).
- [31] Schultz, Matthew, and Thorsten Joachims. Learning a distance metric from relative comparisons. *Advances in neural information processing systems* 16 (2003).
- [32] Brian Kulis. Metric Learning: A Survey. *Foundations and Trends in Machine Learning (FTML)*, 5(4):287-364, 2013.
- [33] Yang, Liu, and Rong Jin. Distance metric learning: A comprehensive survey. *Michigan State University* 2.2 (2006): 4.
- [34] Kilian Q. Weinberger and Lawrence K. Saul. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *J. Mach. Learn. Res.* 10, 207-244. 2012.
- [35] Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. In *Proceedings, CVPR, IEEE*. (2005).
- [36] Kulis, B. Metric Learning: A Survey. *Now Foundations and Trends*. (2013).
- [37] Guo, Jun, et al. "Building hierarchical representations for oracle character and sketch recognition." *IEEE Transactions on Image Processing* 25.1 (2015): 104-118.
- [38] Han, W.; Ren, X.; Lin, H.; Fu, Y.; Xue, X. Self-supervised learning of orc-bert augmentator for recognizing few-shot oracle characters. In *Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November-4 December* (2020).
- [39] Liu, M.; Liu, G.; Liu, Y.; Jiao, Q. Oracle-bone inscription recognition based on deep convolutional neural network. *J. Image Graph*. (2020), 8, 114-119.
- [40] Zhang, Y.K.; Zhang, H.; Liu, Y.G.; Yang, Q.; Liu, C.L. Oracle character recognition by nearest neighbor classification with deep metric learning. In *Proceedings of the 2019 International Conference on Document*

Analysis and Recognition (ICDAR), Sydney, Australia, 20–25 September (2019); pp. 309–314.

- [41] Pascal, A., Planas, A., Lederhos, F., Herrera, N. Extracción de Características utilizando Redes Siamesas y Aumentación aplicada a las Búsquedas por Similitud de Marcas de Ganado. CoNaISI. Entre Ríos, Argentina (2022).