

Mejorando la Identificación de Marcas de Ganado Vacuno: Redes Siamesas en el Aprendizaje de Funciones de Distancia

Federico Stauber¹, Adrián Planas², Andrés Pascal²

^{1,2} GIBD, Facultad Regional Concepción del Uruguay, Universidad Tecnológica Nacional
¹fedestauber@gmail.com, ²{planasa, pascala}@frcu.utn.edu.ar

Abstract. Las Búsquedas por Similitud son importantes en diversas aplicaciones, incluyendo la identificación de marcas de ganado vacuno para el registro ganadero. Para calcular la similitud entre estas marcas, se utilizan funciones de distancia que miden dicha similitud en base a sus características, o en forma directa a partir de las imágenes correspondientes. En esta última década, las Redes Neuronales Profundas Convolucionales (CNN) han alcanzado muy buena performance en el procesamiento de imágenes. En este artículo se propone un método de preprocesamiento, aumentación de datos y modelos de CNN para aprender una función de distancia en un escenario de One-Shot learning utilizando una arquitectura de Redes Siamesas como mecanismo de entrenamiento.

Palabras Clave: Búsquedas por Similitud, Marcas de Ganado, Metric Learning, CNNs, Redes Siamesas, One-Shot Learning, Aumentación.

1 Introducción

Con la incorporación de nuevos tipos de datos como imágenes, audio, video y texto, los modelos de búsqueda tradicionales se vuelven insuficientes. La búsqueda por similitud emerge como una solución para encontrar objetos similares a un elemento de consulta específico en las bases de datos no tradicionales. Los Espacios Métricos [1] son un modelo que formaliza las búsquedas por similitud y permite el uso de métodos de acceso más eficientes.

Este artículo aborda el aprendizaje de funciones de distancia para la búsqueda de similitud de imágenes de marcas de ganado, una tarea esencial para la identificación y registro de nuevas marcas en el Registro Ganadero del Ministerio de Desarrollo Agrario de Buenos Aires. El registro de marcas de ganado está regulado por el Decreto Ley Nacional 22939 - SENASA 1983 [2], que establece en su Artículo III: *“No se admitirá el registro de un diseño de marca igual, ni uno que pueda confundirse con otro, dentro del ámbito territorial de la misma provincia o del territorio nacional. Se incluyen los que representen un diseño idéntico o similar y aquellos en los que uno de los diseños, al superponerse a otro, cubra todas sus partes”*.

Durante el registro de marcas de ganado, la búsqueda por similitud se realiza en forma manual debido a la falta de un sistema automatizado, afectando la eficiencia y eficacia del proceso. La dependencia de la habilidad visual de los empleados y la exhaustividad de la búsqueda se convierte en una limitación significativa.

Este problema presenta tres desafíos importantes para el aprendizaje automático: a) sólo existe una muestra de cada marca (One-Shot Learning), b) la cantidad de marcas está en crecimiento continuo y es muy costoso reentrenar el modelo cada vez que se agrega un nuevo elemento, y c) encontrar una función de distancia que mide correctamente la similitud entre dos imágenes no es una tarea trivial. En este artículo, se presentan modelos de CNN y técnicas de preprocesamiento, aumentación y entrenamiento para creación de una función de distancia que es eficaz en el cálculo del nivel similitud entre imágenes de marcas de ganado.

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

2. Trabajo Relacionado

Esta sección presenta el contexto del presente estudio. Se describen las técnicas claves para abordar el desafío de utilizar la CNN como función de distancia para la comparación por similitud de las marcas de ganado.

2.1 Recuperación de Imágenes Basada en Contenido

La Recuperación de Imágenes Basada en Contenido (CBIR) es un proceso que permite recuperar imágenes de una base de datos considerando ciertas características visuales, como la similitud en colores o formas. Los principales tipos de características utilizados son el color, la textura y la forma [3, 4]. El color se destaca como la característica visual más ampliamente utilizada, en gran parte debido a la facilidad de extraer datos cromáticos de las imágenes [5]. En contraste, obtener datos sobre la forma y la textura [6] es un proceso mucho más complejo y costoso.

Entre las soluciones más populares, los histogramas [7] ocupan un lugar destacado. Estos histogramas son computacionalmente eficientes, pero pueden ser insensibles a cambios en la posición de la cámara, carecen de información espacial y no manejan bien los cambios en el brillo general de la imagen [8].

La forma es una característica visual fundamental para describir el contenido de las imágenes, aunque puede verse afectada por defectos, ruido, oclusión y distorsiones arbitrarias, lo que dificulta su análisis. La forma puede ser descrita por diferentes aspectos [9] tales como: centro de gravedad [10], masa, media, dispersión, varianza, eje de menor inercia, rectangularidad y convexidad. Existen también enfoques más avanzados como el uso de descriptores invariantes: Momentos de Hu, Legendre o Zernike [11, 12, 13], que ofrecen resultados más confiables y precisos.

2.2 Búsquedas en Espacios Métricos

Los CBIR pueden ser generalizados y modelados mediante Espacios Métricos con el fin de lograr búsquedas eficientes. En [1], se muestra 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, el objetivo es recuperar todos los elementos cercanos a un objeto dado, utilizando d como criterio de similitud.

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 :

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

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 estructura de las Redes Neuronales Convolucionales (CNNs) fue propuesta inicialmente por Fukushima en 1988 [16] y durante los años '90, LeCun y otros investigadores [17] desarrollaron algoritmos de aprendizaje basados en el gradiente que mostraron resultados prometedores en la clasificación. Pero no fue hasta esta última década que las CNNs experimentaron mejoras significativas, destacándose en diversos problemas de reconocimiento de patrones.

Algunos ejemplos de las arquitecturas actuales de CNN son AlexNet [18], VGG Net [19] y NiN [20]. Otras alternativas más avanzadas y eficientes han sido propuestas, incluyendo DenseNet [21], GoogLeNet [22], y Residual Networks [23]. Los componentes básicos son casi los mismos para todas las arquitecturas, sin embargo, las diferencias topológicas producen distintos resultados tanto en la eficiencia en el entrenamiento como en la precisión.

Sin embargo, el concepto de similitud entre imágenes no se encuentra inherentemente presente en el algoritmo de aprendizaje de las Redes Neuronales Convolucionales (CNNs). No obstante, recientemente se han desarrollado arquitecturas como las Redes Neuronales Siamesas [24, 25] y Triplet Loss [26], que permiten capturar el concepto de similitud entre imágenes y han demostrado ser eficientes en el reconocimiento de rostros. Las Redes Siamesas consisten en un par de CNNs idénticas que pueden ser utilizadas para extraer vectores de características o

para entrenar funciones de distancia directamente. Durante el entrenamiento, la red ajusta sus parámetros para minimizar la distancia correspondiente a imágenes similares y maximizar la distancia correspondiente a imágenes diferentes.

El uso de Redes Siamesas/CNNs para la extracción de características orientadas a las Búsquedas por Similitud posee dos importantes problemas:

1. El primero es la escasez de instancias disponibles para el entrenamiento, problema conocido como One-Shot Learning o Few-Shots Learning [27, 28]. Esta situación imposibilita el entrenamiento directo de los modelos. La mejor estrategia hasta el momento para abordar este problema es el Transfer Learning/Fine Tuning [29, 30], sin embargo, esta técnica se ve limitada por la disponibilidad de bases de datos de características similares.
2. El segundo desafío es el de la generalización: cuando se entrena una CNN para extraer vectores característicos, el comportamiento del modelo suele estar fuertemente ligado a las clases con las cuales se realizó el entrenamiento. Por esta razón no logra adaptarse de manera efectiva a nuevas imágenes que se incorporen a la base de datos si no se la re-entrena. Una alternativa estudiada recientemente es el Aprendizaje Métrico (Metric Learning) [31], que propone aprender directamente una función de distancia correspondiente a los datos existentes en lugar de extraer características. Este es el enfoque utilizado en este estudio.

2.4 Aprendizaje de Funciones de Distancia (Metric Learning)

El aprendizaje métrico [32] se ocupa del aprendizaje de funciones de distancia que se ajustan a una determinada definición de similitud. En problemas de reconocimiento de patrones, la definición de similitud es específica a la tarea y el éxito del aprendizaje depende de la alineación del aprendizaje con la tarea en cuestión.

En [33] se realiza un análisis profundo de los métodos de aprendizaje de métricas y se da una visión general de las diversas técnicas y enfoques aplicados en el campo. En [34] se plantea un enfoque de aprendizaje métrico basado en márgenes. En otro trabajo relacionado, se describe una técnica de aprendizaje de métricas discriminativas aplicada al desafío de la verificación de rostros [35]. En [36] se proporciona una visión general del aprendizaje de métricas, cubriendo diferentes enfoques y aplicaciones y destacando los desafíos y tendencias en el campo.

2.5 Búsquedas por Similitud de Marcas de Ganado

En investigaciones previas se han usado varios enfoques sobre este tema. En [37] se emplearon histogramas de tangentes para cada punto de la imagen y se compararon utilizando el coeficiente de Pearson; otra propuesta consistió en el uso de histogramas generados midiendo distancias entre pares aleatorios de puntos de la imagen [38].

En un enfoque más prometedor, se obtuvieron mejores resultados en [39] utilizando Momentos de Hu y Legendre. En [40], se aplicó una CNN preentrenada para extraer características de las marcas de ganado, seguida de SVM para clasificación. Aunque alcanzaron una tasa de reconocimiento del 93%, persiste el desafío de agregar nuevas marcas a la base de datos, ya que éste método implica reentrenar el SVM cada vez que se incorporaba una nueva marca. Una solución más

prometedora se presenta en [41], donde se utiliza una técnica llamada ALR3, basada en histogramas de propiedades geométricas de las imágenes. Aunque muestra potencial, aún requiere una evaluación más extensa. En un estudio anterior [42], presentamos el uso de una CNN entrenada mediante Redes Siamesas y técnicas estándares de Aumentación de datos, junto con una técnica propia para mejorar la eficacia de las Búsquedas por Similitud de Marcas de Ganado.

3 Aprendizaje de la Función de Distancia

En esta sección se describe la base de datos y los modelos utilizados para el aprendizaje de funciones de distancia a partir de una base de datos compuesta por sólo una instancia de cada marca, utilizando Redes Siamesas, técnicas clásicas de Aumentación y el método de relieve presentado en [42].

3.1 Base de Datos de Marcas

El conjunto de datos utilizado consta de 203 marcas obtenidas del Registro Ganadero del Ministerio de Desarrollo Agrario de Buenos Aires. La Fig. 1 exhibe algunos ejemplos de estas marcas.

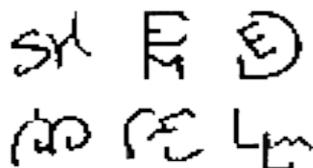


Fig. 1. Marcas de ganado pertenecientes a la base de datos.

3.2 Modelos utilizados

En este estudio se probaron doce modelos, de los cuales los tres mejores (Mod_1 , Mod_2 y Mod_3) se muestran en este artículo. Los mismos consisten en CNNs de cinco (Mod_3) y seis (Mod_1 y Mod_2) capas convolucionales, todas con activación ReLU. Cada capa convolucional está intercalada con una capa de Batch-Normalization. A pesar de estas configuraciones en común, estos modelos se diferencian por las siguientes características:

- Mod_1 : 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. Se emplearon kernels de (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.
- Mod_2 : Este modelo utiliza un tamaño de vector de salida de 256 neuronas. Las cuatro primeras capas convolucionales emplean 32 kernels, mientras que las

restantes utilizan 64. Todos los kernels son de (3, 3). Adicionalmente, se incorporan una capa de MaxPooling2D y dos capas de Dropout. Luego se incluye una capa "flatten" antes de pasar a una capa densa de activación softmax, que representa la salida utilizada por la red siamesa. Este modelo cuenta con 664.080 parámetros entrenables.

- *Mod₃*: En este modelo, se utiliza la propiedad padding='same' en las capas de convolución para mantener las dimensiones de la imagen. Además posee 5 capas convolucionales en lugar de 6, como los modelos anteriores. Las cuatro primeras capas convolucionales se intercalan entre 32 y 64 kernels, mientras que la restante es de 256. Los kernels son de (3, 3) para todas las capas convolucionales. Adicionalmente, incorpora una capa MaxPooling2D, dos capas de Dropout y una capa GlobalAveragePooling2D. Finalmente, el modelo incluye una capa "flatten" y una capa densa de 128 neuronas que representa la salida utilizada por la red siamesa. Contiene 237.280 parámetros entrenables.

A cada uno de los modelos presentados se lo utilizó 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 son de 28x28 píxeles.

3.3 Aumentación de Datos

En el ámbito de Búsquedas por Similitud, es común disponer de una sola muestra para cada objeto de la base de datos. Este hecho plantea un desafío significativo al emplear Redes Neuronales Convolucionales (CNNs) como método para determinar distancia de similitud, debido a las limitaciones mencionadas previamente. En el presente estudio, se solucionó este problema mediante la generación de 91.350 muestras adicionales (450 por cada marca) a partir de las 203 imágenes originales de la base de datos. Estas muestras se obtuvieron al azar, en principio mediante técnicas estándar de aumento de datos para imágenes en blanco y negro:

- Traslación: en direcciones este, oeste, norte y sur, hasta un 25% de la posición de la imagen original.
- Rotación: en sentido horario y antihorario, hasta un ángulo de 45 grados.
- Escalado: entre un 70% y un 130% del tamaño original.
- Cortes: hasta un 30% de la cantidad de píxeles de la imagen.
- Ruido Gaussiano: agregando hasta un 20% más de pixels en la imagen.

Y posteriormente se les agregó el procesamiento "Relieve", propuesto en [42], que consiste en aplicar una expansión linealmente decreciente hacia los lados de las líneas que componen la imagen. Esta técnica produce mejoras significativas en la eficacia del entrenamiento, según nuestros experimentos. La Fig. 2 exhibe dos marcas (izquierda) después de su preprocesamiento y sus correspondientes relieves (derecha).

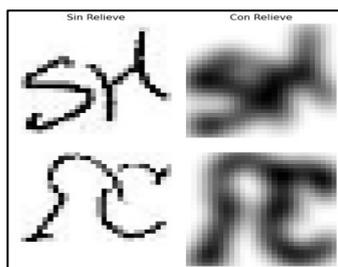


Fig. 2. Marcas y sus Relieves.

Los resultados experimentales demuestran que este mecanismo contribuye significativamente a que la Red Siamesa realice predicciones más robustas y estables en comparación con aquellas basadas exclusivamente en la aumentación estándar.

4 Experimentos Realizados

Para el entrenamiento de la Red Siamesa se seleccionaron al azar 406.000 pares de imágenes aumentadas, asegurando un equilibrio entre pares similares y no similares y garantizando la participación de cada marca en al menos 2.000 pares.

Se llevaron a cabo pruebas para evaluar la eficacia de los modelos mediante la función de distancia obtenida a partir de la Red Siamesa y los mecanismos de aumentación. Estas pruebas involucraron dos lotes de consultas diferentes, los cuales fueron:

- QD: 40 imágenes de consulta dibujadas a mano, las cuales se crearon copiando marcas existentes en la base de datos. Un total de 20 consultas se dibujaron en papel y posteriormente se digitalizaron, mientras que las otras 20 se dibujaron utilizando una aplicación de diseño gráfico.
- QN: para evaluar la capacidad de generalización de la función de distancia sin reentrenamiento, se incorporaron 10 nuevas imágenes a la base de datos y se generaron 10 consultas dibujadas a mano, una por cada nueva marca. La Red Siamesa no se reentrenó con las nuevas marcas. Este experimento tuvo como objetivo determinar si este método puede manejar nuevos elementos (nuevas "clases") de manera suficientemente efectiva sin requerir reentrenamiento.

Los modelos fueron entrenados en bloques de 5 épocas con 406.000 pares distintos de imágenes. En total se realizaron 150 épocas de entrenamiento alcanzando los valores 0,969, 0,983 y 0,995 de precisión para los modelos 2, 3 y 1 respectivamente. 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 adecuada en el contexto del problema real planteado. Los resultados se compararon con el trabajo previo de extracción de características mediante redes siamesas y uso de la distancia euclidiana como función de distancia, aplicado a la misma base de datos y lotes de consulta [42]. Este modelo anterior lo denominamos Mod₀ y lo usamos como base de comparación.

5 Resultados y Discusión

En la Tabla 1 se muestran los porcentajes de acierto en las búsquedas por similitud para los k vecinos mas cercanos (NN_k), para $k=1, 3$ y 5 . Como podemos observar, los modelos Mod_1 y Mod_3 superan al modelo anterior Mod_0 en las búsquedas de las imágenes dibujadas (QD) de las marcas con las que fueron entrenados. De estos dos, Mod_1 es el que obtiene mejores resultados. Sin embargo, ante el alta de nuevas marcas y búsquedas sin reentrenamiento (QN), Mod_3 resulta más consistente, con mayor poder de generalización, y supera significativamente al modelo anterior (90% para el NN_5 vs 60%).

Tabla 1. Porcentajes de Acierto de los distintos modelos.

Experimento	NN_1	NN_3	NN_5
Mod_0 -QD	-	82,50	85,00
Mod_1 -QD	90,00	93,00	97,00
Mod_2 -QD	55,00	72,00	75,00
Mod_3 -QD	80,00	93,00	95,00
Mod_0 -QN	-	50,00	60,00
Mod_1 -QN	10,00	50,00	80,00
Mod_2 -QN	30,00	40,00	60,00
Mod_3 -QN	20,00	70,00	90,00

Una característica importante de estas CNNs es que a pesar de ser profundas son relativamente pequeñas (Mod_3 posee solo 237.280 parámetros), lo que permite que su uso como función de distancia no sea excesivamente costoso (en los experimentos, cada ejecución de la función de distancia se realizó en alrededor de 0,03 segundos).

A pesar de ello, hay que notar que estos modelos no cumplen con las propiedades de las distancias métricas. De hecho, la función obtenida no es reflexiva, ni simétrica y tampoco cumple la desigualdad triangular. La simetría se resuelve fácilmente haciendo $d(x,y) = CNN(x,y) + CNN(y,x)$. 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 el cumplimiento de esta propiedad, alcanzando un 98,08% de éxito para Mod_3 . 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).

6 Conclusiones y Trabajo Futuro

En este artículo se presentaron modelos de CNN para el de Redes Siamesas, con Aumentación estándar de los datos y una técnica propia denominada Relieve para obtener una función distancia a utilizar en la búsqueda de imágenes de marcas de ganado ante una situación de One-Shot Learning. Como parte de este estudio se

presentaron tres modelos de CNN, de los cuales uno de ellos alcanza niveles de acierto suficientemente altos como para ser implementado en un sistema real.

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

- Realizar experimentos sobre la BD completa (alrededor de 60.000 marcas).
- Probar el uso de índices métricos con Mod₃, para determinar su performance y el porcentaje real de elementos descartados erróneamente.
- Analizar estrategias para que la red generalice mejor, de tal manera de que sea robusta ante nuevas imágenes sin reentrenamiento.

Referencias

1. Chávez, Edgar, et al. Searching in metric spaces. *ACM computing surveys (CSUR)* 33.3: 273-321 (2001).
2. Decreto Ley Nacional 22939 SENASA. Servicio Nacional de Sanidad y Calidad Agroalimentaria. Título I - De las marcas y señales en general. Artículos 1 al 4. (1983).
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. Sampallo, Guillermo M., et al. Reconocimiento de marcas de ganado. IX Congreso Argentino de Ciencias de la Computación (2003).
38. Sánchez Torres, Germán, and Manuel E. Rodríguez García. Medida de similitud entre imágenes de marcas de ganado mediante distribuciones de forma. Revista Ingenierías Universidad de Medellín 13.25: 177-189 (2014).
39. Sanchez, G., and M. Rodriguez. Cattle marks recognition by hu and legendre invariant moments. ARPN Journal of Engineering and Applied Sciences 11.1: 607-614 (2016)
40. Silva, Carlos, et al. Cattle Brand Recognition using Convolutional Neural Network and Support Vector Machines. IEEE Latin America Transactions 15.2: 310-316 (2017).
41. Pascal, A., Michel, L., Romani, R. Zernike Moments vs ALR3 Applied to Similarity Searching of Cattle Brands. VI SABTIC (2018).
42. 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. CoNaIISI. Entre Ríos, Argentina (2022).