

Extracción de Características utilizando Redes Siamesas y Aumentación aplicada a las Búsquedas por Similitud de Marcas de Ganado

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 Lederhos

*FRCU-Universidad Tecnológica Nacional
fedelederhos5@gmail.com*

Norma E. Herrera

*Universidad Nacional de San Luis
nherrera@unsl.edu.ar*

Resumen

Las Búsquedas por Similitud constituyen un importante campo de estudio en el presente. Un ejemplo de su aplicación es la búsqueda de marcas de ganado, necesaria ante la solicitud de alta de una nueva marca al Departamento de Registro Ganadero. Para calcular su similitud, se suelen utilizar funciones de distancia métrica aplicadas a vectores de características extraídas a partir de su imagen. Existen varios métodos de extracción de características, a los cuales en la última década se le han sumado las Redes Neuronales Profundas Convolucionales (CNN). En este artículo se muestra el uso de una CNN entrenada mediante Redes Siamesas y con técnicas estándares de Aumentación de datos, a las cuales se le añade una técnica propia para mejorar la eficacia de la extracción de características aplicadas a las Búsquedas por Similitud de Marcas de Ganado. Además, se evalúa experimentalmente su performance.

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

1. Introducción

Las bases de datos tradicionales se construyen en torno al concepto de datos estructurados y consultas exactas: la base de datos se divide en registros, cada registro comprende claves totalmente comparables. Consultar la base de datos devuelve todos los registros cuyas claves son exactamente iguales al valor proporcionado. Actualmente, las bases de datos poseen también la capacidad de almacenar nuevos tipos de datos como imágenes, audio, video y texto. Por lo tanto, los modelos de búsqueda tradicionales ya no son útiles dentro de estos nuevos marcos, principalmente debido a que los datos no están estructurados; por lo que no siempre es posible organizarlos en registros y campos,

haciendo inútiles las búsquedas exactas. La búsqueda por similitud proporciona una forma de encontrar objetos de la base de datos que sean similares a un elemento de consulta determinado. Los Espacios Métricos [1] constituyen un modelo que formaliza el concepto de búsqueda por similitud en bases de datos no tradicionales y que permite utilizar métodos de acceso que mejoran la eficiencia de la búsqueda.

En este artículo nos enfocamos en la búsqueda de similitud de imágenes de marcas de ganado, necesarias para identificar cada nueva marca a ser registrada 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”.

El proceso de registro de la marca se inicia llevando la imagen de la marca a las autoridades del Registro Ganadero. Luego se realiza una búsqueda para establecer que no se han registrado marcas similares previamente. Actualmente no existe un sistema computacional que permita este tipo de búsquedas, por lo que este proceso se realiza mediante registros físicos. Esto disminuye la eficiencia y la eficacia del proceso, que dependen en gran medida de la capacidad visual de los empleados y la completitud de la búsqueda.

En este artículo se experimenta la aplicación de Redes Siamesas y Aumentación como método de extracción de características para la búsqueda de imágenes de marcas de ganado y se propone un mecanismo de transformación de las imágenes aumentadas llamado “Relieve”, que produce mejoras significativas en la extracción de características.

El resto de este documento está organizado de la siguiente manera: la Sección 2 presenta el trabajo relacionado, incluida una breve explicación de la

búsqueda de imágenes por contenido, CNNs, Redes Siamesas, el modelo de espacio métrico y la búsqueda de marcas de ganado. En la Sección 3, se describe la estructura genérica del Sistema de Búsqueda. El proceso de extracción de características y el mecanismo de Relieve se explican en detalle en la Sección 4. La Sección 5 describe los experimentos realizados y la Sección 6 los resultados obtenidos. Finalmente, en la Sección 7 se presentan las conclusiones y el trabajo futuro.

2. Trabajo Relacionado

En esta sección se describe el contexto de este estudio: Búsquedas de Imágenes por Contenido aplicada a las Consultas por Similitud de Marcas de Ganado, y las técnicas utilizadas (CNNs, Redes Siamesas, Aumentación de datos) para resolver el problema de la extracción de características a partir de imágenes para su posterior comparación por similitud.

2.1. Recuperación de Imágenes Basada en Contenido

La recuperación de imágenes basada en contenido (CBIR) es el proceso de recuperación de imágenes de una base de datos teniendo en cuenta algunas características visuales de esas imágenes, por ejemplo, la búsqueda de imágenes con contenidos similares de colores o formas. CBIR funciona principalmente extrayendo características de la imagen de consulta y buscando estas características en la base de datos. Esas características generalmente están representadas por vectores tanto para la imagen de consulta como para las imágenes de la base de datos. Hay tres tipos principales de características: color, textura y forma [3, 4].

Idealmente, estas características deberían integrarse para proporcionar una mejor discriminación en el proceso de comparación. El color es, con mucho, la característica visual más común utilizada en CBIR, principalmente debido a la simplicidad de extraer datos de color de las imágenes [5]. Extraer información sobre forma y textura [6] son tareas mucho más complejas y costosas.

Los histogramas [7] constituyen una de las soluciones más populares para modelar características de imágenes. Cada histograma describe un nivel de gris o una distribución de color para una imagen dada y son computacionalmente eficientes, pero generalmente insensibles a pequeños cambios en la posición de la cámara. Los histogramas de color también tienen algunas limitaciones. Un histograma de color no proporciona información espacial; solo describe qué colores hay en la imagen y en qué cantidades. Además, los histogramas de color también son sensibles a los cambios en el brillo general de la imagen [8].

La forma es una característica visual importante y es una de las características básicas utilizadas para

describir el contenido de la imagen. Sin embargo, la representación y descripción de formas es una tarea difícil. La forma puede corromperse con defectos, ruido, oclusión y distorsión arbitraria. Además de esto, no se sabe qué características son más importantes en la forma. Básicamente, la recuperación de imágenes basada en formas consiste en medir la similitud entre las formas representadas por sus características. Algunas características geométricas simples se pueden usar para describirlas. Por lo general, esas características geométricas solo pueden discriminar formas con grandes diferencias y no son adecuadas para descriptores de forma independientes. 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. Un mejor enfoque para la representación de formas es usar descriptores invariantes como Momentos de Hu, Legendre o Zernike [11, 12, 13].

2.2. Redes Neuronales Convolucionales (DCNN o CNN)

Esta estructura fue propuesta originalmente por Fukushima en 1988 [14]. Inicialmente no fue muy utilizada por limitaciones de hardware para el entrenamiento de la red. Durante los '90, LeCun y otros [15] aplicaron un algoritmo de aprendizaje basado en el gradiente y obtuvieron buenos resultados en la clasificación de dígitos manuscritos. Posteriormente se mejoraron notablemente las CNNs, alcanzando buenos resultados en muchos otros problemas de reconocimiento. Las CNNs poseen importantes ventajas sobre las DNNs originales, en particular, el poseer un mecanismo más parecido a la forma en que un humano visualiza una imagen, la optimización para el procesamiento de imágenes en 2D y 3D, y la efectividad en la extracción de características en 2D. Las capas de submuestreo máximo (Max Pooling) que poseen, son efectivas en la absorción de pequeñas variaciones de forma. Además, reducen significativamente la cantidad de parámetros a entrenar, en comparación con una Red Neuronal Densa (Fully Connected) de tamaño similar. La arquitectura general de una CNN consiste en dos partes principales: Extracción de Características y Clasificador. La extracción de características se realiza mediante capas de Convolución y de SubMuestreo (Pooling), mientras que la clasificación usualmente se lleva a cabo a través de capas densas, que obtienen mejores resultados para esta tarea [16, 17].

Las arquitecturas actuales de CNNs consisten típicamente en la combinación de varias capas convolucionales y de pooling, en su mayoría con activación ReLU, seguidas por capas densas más SoftMax hacia el final. Algunos ejemplos de tales modelos son LeNet [15], AlexNet [18], VGG Net [19], NiN [20] y All Conv [21]. Otras alternativas más avanzadas y eficientes han sido propuestas, incluyendo DenseNet [22], FractalNet [23], GoogLeNet [24, 25,

26], y Residual Networks [27]. 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 en la clasificación.

El concepto de similitud entre imágenes no está reflejado directamente en el algoritmo de aprendizaje de las CNNs. Recientemente las arquitecturas de Redes Neuronales Siamesas [28, 29] y Triplet Loss [30, 31, 32] han sido desarrolladas para la extracción de características que capturan el concepto de similitud entre imágenes de entrada, y se han utilizado eficientemente en el reconocimiento de rostros. Las Redes Siamesas poseen dos o más CNNs idénticas que se utilizan para extraer vectores característicos que luego son comparados a través de una función de distancia. Durante el entrenamiento, la red modifica sus parámetros de tal manera de minimizar la distancia entre dos vectores correspondientes a imágenes similares, y maximizar la distancia cuando las imágenes son diferentes.

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

1. El primero es que, típicamente se cuenta con solo una (o unas pocas) instancias de cada objeto, lo que imposibilita el entrenamiento directo. Este problema es conocido como One-Shot Learning o Few-Shots Learning [33, 34] y actualmente su mejor estrategia de solución se basa en el Transfer Learning [35-38]. Cuando se cuenta con modelos ya entrenados sobre una base de objetos con características muy similares, los resultados son muy buenos. Pero esta técnica se ve limitada a la existencia de dichas bases de datos. En algunos trabajos recientes como [39] se proponen algoritmos que podrían superar estas limitaciones.
2. El segundo problema es que los vectores resultantes de la extracción de características basadas en CNNs (y de la mayoría de los métodos de aprendizaje automático actuales) están fuertemente asociados a las clases con las cuales se entrenan los modelos, y no generalizan suficientemente bien. Es decir, no funcionan con la misma eficacia para las nuevas imágenes que se incorporan a la base de datos (imágenes que corresponderían a nuevas “clases”). Para ello una alternativa que se ha estudiado en los últimos años es el Aprendizaje Métrico (Metric Learning) [40, 41], que en lugar de extraer características propone directamente aprender la función de distancia correspondiente a los datos existentes. Esta alternativa quedó fuera del alcance de este artículo.

2.3. Búsquedas en Espacios Métricos

Los CBIR se pueden generalizar y modelar mediante Espacios Métricos con el objetivo de que las búsquedas sean eficientes. En [1] se muestra que el problema de

búsqueda de similitud se puede expresar de la siguiente manera: dado un conjunto U de objetos y una función de distancia d definida entre ellos que cuantifica su similitud, el objetivo es recuperar todos los elementos similares a un objeto dado usando d como criterio.

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)

Cuanto menor es la distancia entre dos objetos, más similares son. El par (U, d) se llama espacio métrico. Un subconjunto finito X de U , al que llamaremos base de datos, es el conjunto de objetos sobre el cual se está realizando la búsqueda. Hay básicamente dos tipos de consultas por similitud que son interesantes en espacios métricos:

- (a) *Consulta por Rango o $(q, r)_d$* : devuelve todos los elementos que se encuentra como máximo a una distancia r de q .

$$(q, r)_d = \{x \in X / d(q, x) \leq r\}$$

- (b) *Consulta de los k Vecinos Más Cercanos o $NN_k(q)_d$* : recupera los k elementos de X , más cercanos a q .

$$NN_k(q)_d = A,$$

$$|A| = k,$$

$$A = \{x \in X / \forall y \in (X - A), d(q, x) \leq d(q, y)\}$$

Dada una base de datos de n objetos, estas consultas pueden responderse de forma trivial realizando n evaluaciones de distancia. Desafortunadamente, esto es generalmente muy costoso en aplicaciones reales. La importancia de modelar estas consultas a través de espacios métricos radica en la existencia de índices que utilizando la desigualdad triangular descartan elementos de la base de datos sin necesidad de compararlos con la consulta, haciendo mucho más eficiente la búsqueda [42-46].

2.4. Búsquedas por Similitud de Marcas de Ganado

Existen precedentes en la investigación de marcas de ganado por similitud. En [47] se utilizan histogramas de pendientes de tangentes para cada punto de la imagen y luego se comparan utilizando el coeficiente de Pearson. El problema de usar tangentes es que son muy sensibles a la rotación, incluso en pequeñas variaciones. Además, los resultados mostrados no son concluyentes. Otra línea de investigación propuesta utiliza histogramas generados midiendo distancias entre pares aleatorios de puntos de la imagen y evalúa la similitud utilizando distancias de Minkowski [48]. En nuestras pruebas, este enfoque no obtuvo buenos resultados. Se muestran mejores resultados en [49] usando Momentos de Hu y

Legendre con una tasa de éxito similar a la de los Momentos de Zernike. En [50] se utiliza una CNN preentrenada para extraer características de marcas de ganado y luego SVM como mecanismo de clasificación. En el estudio se utiliza una base de datos de 12 marcas con 45 muestras de cada una. Si bien la tasa de reconocimiento alcanzada es de alrededor del 93%, persiste el problema de agregar nuevas marcas a la base de datos, ya que implica que habría que reentrenar el SVM cada vez que se da un alta. Además, este método de clasificación no es eficiente para numerosas clases (más de 60.000 en el problema real planteado). Una mejor solución se presenta en [51], en la cual se utiliza una técnica especialmente diseñada para la extracción de características de marcas, denominada ALR3, basada en histogramas de propiedades geométricas de las imágenes, aunque aún no ha sido suficientemente probada.

3. Estructura del Sistema de Búsquedas de Marcas por Similitud

En un sistema de búsquedas por similitud de marcas, las imágenes se almacenan en la base de datos junto con sus vectores de características. Para calcular los vectores, primero se preprocesa cada imagen para limpiarla y luego se ejecuta un algoritmo para extraer sus características. Se crea un índice métrico tomando como entrada los vectores de características para que el proceso de búsqueda sea más eficiente. A continuación se resumen los pasos de preparación de la BD (inserción de imágenes) y los correspondientes a las consultas.

Inserción de Marcas en la BD: cada vez que se da de alta a un nuevo elemento se ejecutan los siguientes pasos:

1. Preprocesamiento de la Imagen: este paso incluye procesos tales como binarización, esqueletización, obtención del rectángulo delimitador mínimo (MBR), redimensionamiento de la imagen a una medida estándar, reducción de ruido y normalización.
2. Extracción de Características: para cada nueva imagen, se crea un vector de características. En este artículo se estudia la extracción utilizando una CNN entrenada a través de una Red Siamesa. Para esto, en forma previa se debe realizar el entrenamiento de la misma.
3. Almacenamiento de la imagen y junto a su vector correspondiente en la base de datos.
4. Actualización del índice métrico mediante la inserción del nuevo elemento.

Consulta por Similitud: dada una imagen de consulta, se realiza una Búsqueda por Rango o de los k Vecinos más Cercanos (NN_k) de la siguiente manera:

1. Preprocesamiento de la Consulta.

2. Extracción de Características.
3. Uso del índice métrico y la función de distancia (en este caso Manhattan) para descartar elementos sin compararlos con el vector de características de la consulta y así devolver el conjunto resultante de imágenes similares.

Es de destacar, que no es factible operativamente el reentrenamiento de la Red ante el alta de cada nueva imagen, por lo que la extracción de características debe funcionar correctamente tanto para imágenes existentes en la BD al momento del entrenamiento como para nuevas imágenes.

4. Extracción de Características mediante Redes Siamesas y Aumentación

En esta sección se describe el mecanismo utilizado para realizar la extracción de características a partir de una base de datos compuesta por sólo una instancia de cada marca, utilizando Redes Siamesas, técnicas clásicas de Aumentación y un método propio presentado como parte de este trabajo.

4.1. Base de Datos de Marcas

EL conjunto de datos utilizado está compuesto por 203 marcas extraídas de la base de datos del Registro Ganadero del Ministerio de Desarrollo Agrario de Buenos Aires. En la Figura 1 se muestran algunos ejemplos de dichas marcas.

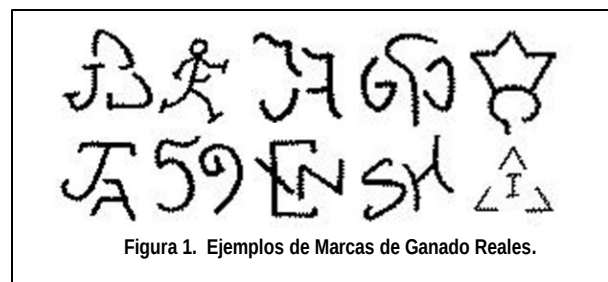


Figura 1. Ejemplos de Marcas de Ganado Reales.

Las imágenes son bastante diferentes en términos de diseño; algunas están compuestas por números, letras o combinación de éstos, y otras poseen una forma totalmente arbitraria.

4.2. Red Siamesa Utilizada

El modelo utilizado es una Red Neuronal Convolutiva Siamesa con seis capas convolucionales alternadas mediante capas de Max-Pooling y Batch-Normalization. Las tres primeras capas convolucionales utilizan 32 kernels y las demás 64. Cuatro de ellas tienen kernels de (3, 3) y las otras dos de (5, 5). Además hay dos capas Dropout para evitar overfitting. Luego de

las convoluciones la red posee una capa Flatten y por último una capa Densa (Fully-Connected). Se utilizó la función de activación ReLU en todas las capas convolucionales. Los vectores obtenidos de esta manera a partir de dos imágenes de entrada, se comparan mediante la distancia de manhattan (L_1) y luego se aplica la función Sigmoide como activación.

Como función de pérdida se utilizó Binary Cross Entropy y el optimizador fue Adam. El modelo posee 324.864 parámetros entrenables, con imágenes de 28x28 pixels como entrada.

Los vectores resultantes de la capa Densa son los que luego se utilizan durante las búsquedas por similitud como vectores característicos y se comparan mediante la distancia de manhattan (L_1) para determinar su similitud.

4.3. Aumentación Estándar

En los problemas reales de Búsquedas por Similitud se suele contar con una sola muestra de cada objeto de la base de datos utilizada. Esto es un gran problema cuando se decide utilizar CNNs como método de extracción de características por las razones anteriormente nombradas.

En este caso de estudio, a partir de 203 imágenes originales en la base de datos, se generaron 203.000 muestras (1.000 por cada marca) mediante las siguientes técnicas 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 pixels de la imagen.
- Ruido Gaussiano: agregando hasta un 20% más de la cantidad de pixels de la imagen.

Posteriormente, tomando como base las imágenes aumentadas, se obtuvieron en forma aleatoria 203.000 pares, balanceando las cantidades de pares similares y no similares, y asegurando que cada marca participe en al menos 1.000 pares. Estos pares se utilizaron para el entrenamiento de la Red Siamesa.

4.4. Cálculo del Relieve

Como se verá más adelante en los experimentos, el entrenamiento con la aumentación estándar utilizada no produjo buenos resultados, salvo para el caso en de las consultas generadas de la misma manera. Por esta razón se diseñó un procedimiento complementario de aumentación, que es robusto ante pequeños cambios en

la posición de los pixels de las imágenes. Este procedimiento consiste en una expansión linealmente decreciente hacia los lados, de las líneas que constituyen la imagen. A esta transformación la denominamos “Relieve”.

```

Function Relieve(img, radio)
  newimg = zeros_like(img)
  foreach black point(x,y) in img:
    for i in 1..(radio*2):
      for j in 1..(radio*2):
        xa = x+radio-i
        ya = y+radio-j
        if (xa in 1..xmax)&(ya in 1..ymax):
          dist = sqrt((xa-x)**2 +
                    (ya-y)**2)
          if dist<=radio:
            newimg[xa,ya]+=abs(radio-dist)
  return newimg
  
```

Figura 2. Pseudocódigo del Relieve.

El algoritmo de cálculo del relieve se presenta la Figura 2. Como parámetros se ingresa la imagen y el radio de expansión. En este caso utilizamos radio 5. En la Figura 3 se muestran dos marcas (izquierda) preprocesadas y sus respectivos relieves (derecha).

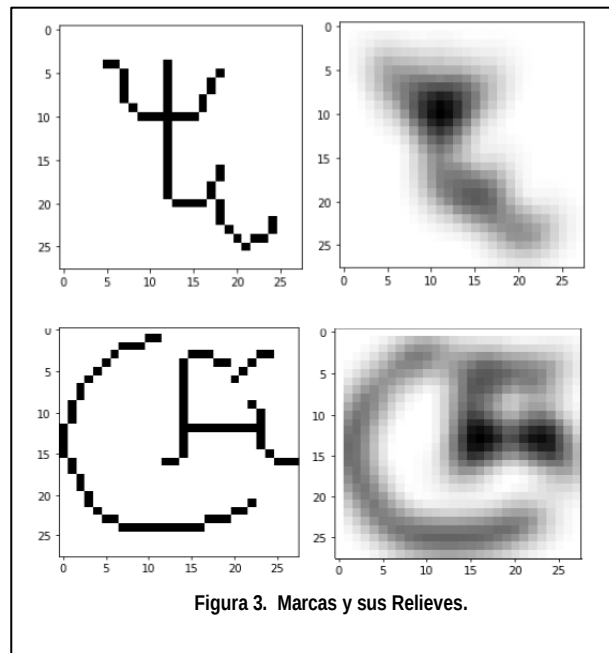


Figura 3. Marcas y sus Relieves.

Como parte del proceso de aumentación, se calcularon los relieves de todos los pares generados por aumentación estándar y luego fueron utilizados para el entrenamiento de la Red Siamesa. Durante las búsquedas por similitud, se calcularon los relieves de cada imagen de consulta para que la extracción de sus características siga el mismo procedimiento.

Según los experimentos, este mecanismo ayuda significativamente a la Red Siamesa a realizar una extracción de características que es mucho más robusta y estable que la basada solo en aumentación estándar.

5. Experimentos Realizados

Para verificar la eficacia de la extracción de características mediante la Red Siamesa y los mecanismos de aumentación, se realizaron distintas pruebas combinando dos mecanismos de aumentación (con y sin relieve) con tres lotes de consultas distintos.

Los lotes de consulta fueron:

- 100 consultas generadas aleatoriamente mediante aumentación estándar, es decir, de la misma manera en que se generaron las muestras de entrenamiento.
- 40 consultas dibujadas a mano, copiando marcas existentes en la BD por distintas personas. 20 de ellas se dibujaron en papel y luego se digitalizaron y el resto se dibujó utilizando una aplicación de diseño gráfico.
- 10 marcas nuevas. En este caso se incorporaron a la BD 10 nuevas imágenes y se generaron otras 10 Consultas dibujadas a mano, una por cada nueva marca. La Red Siamesa no se reentrenó con las nuevas marcas. Este experimento se realizó específicamente para conocer si estos métodos de extracción de características generalizan suficientemente bien como para manejar nuevos elementos (nuevas “clases”) sin reentrenamiento.

En base a estos dos métodos de extracción de características y los tres lotes de consultas, se realizaron los siguientes experimentos:

- SinRelQA: extracción de características con aumentación estándar y búsquedas de 100 consultas generadas aleatoriamente a través de aumentación.
- SinRelQD: extracción de características con aumentación estándar y búsquedas de 40 consultas dibujadas a mano, copiadas a partir de imágenes de la BD.
- SinRelQN: extracción de características con aumentación estándar; alta de 10 nuevas marcas en la BD y búsquedas de 10 consultas dibujadas a mano, una por cada imagen nueva.
- ConRelQD: extracción de características con aumentación estándar más relieve, y búsquedas de 40 consultas dibujadas a mano, copiadas a partir de las imágenes de la BD.
- ConRelQN: extracción de características con aumentación estándar más relieve; alta de 10 nuevas marcas en la BD y búsquedas de 10

consultas dibujadas a mano, una por cada imagen nueva.

Previo a las pruebas, se asoció a cada consulta la marca de la base de datos que debía devolverse como similar y luego se calcularon los 3 y 5 vecinos más cercanos (NN_3 , NN_5) para cada consulta.

El entrenamiento de la Red Siamesa se realizó en 30 épocas para ambos casos (con y sin relieve), alcanzando una precisión de 0.9766 y pérdida de 0.0434, con una precisión de validación igual a 0.9836 para el lote de aumentación sin relieve, y precisión 0.9865, pérdida 0.0313 y precisión de validación 0.9884 para el entrenamiento con relieve. Cada entrenamiento se realizó en aproximadamente una hora en una PC con procesador i5, 16 GB RAM y GPU GeForce GTX 960 con 1024 núcleos CUDA.

Se utilizó la tasa de aciertos (porcentaje) para medir la eficacia de estos experimentos ya que es el indicador que mejor representa si la solución al problema real planteado es adecuada o no.

6. Resultados y Discusión

Como era de esperar, el porcentaje de acierto para las consultas generadas mediante el mismo mecanismo de aumentación que fue utilizado para el entrenamiento de la red, es significativamente alto (99,00% para los NN_5 , ver Tabla 1). Esto es debido a que las CNNs brindan excelente resultados cuando el lote de entrenamiento es representativo de todos los casos de consultas.

En el problema real, las consultas por similitud de las marcas de ganado son imágenes dibujadas mediante alguna aplicación gráfica o inclusive en papel. Ante estos lotes de consulta, utilizando sólo aumentación estándar, los porcentajes disminuyen notablemente: 32,50% para los 5 vecinos más cercanos de las consultas dibujadas y sólo un 20,00% de aciertos para las consultas nuevas.

Por otro lado, cuando se utilizó el Relieve durante el entrenamiento y en las consultas, los aciertos aumentaron considerablemente, alcanzando un 85,00% para las consultas dibujadas y un 60,00% para las consultas nuevas, es decir, valores significativamente superiores que los obtenidos con aumentación estándar.

Tabla 1. Porcentajes de Acierto

Experimento	NN_3	NN_5
SinRelQA	94,00%	99,00%
SinRelQD	25,00%	32,50%
SinRelQN	20,00%	20,00%
ConRelQD	82,50%	85,00%
ConRelQN	50,00%	60,00%

Estos resultados, si bien aún no son lo suficientemente buenos para implementarlos en una

aplicación real, son prometedores y pueden mejorarse realizando algunas modificaciones. El mayor problema de la Red Siamesa/CNN, es la falta de generalización adecuada para manejar casos que no formaron parte de su entrenamiento.

Analizando las consultas ConRelQD que fallaron, se nota que una diferencia importante de aspecto en las proporciones entre la consulta y el elemento buscado produce resultados erróneos, tal como se ve en los dos primeros elementos de la Tabla 2. En el primer caso la consulta es mas alta y en el segundo, mas delgada. Lo mismo sucede con diferencias en las posiciones relativas de las partes que componen una imagen, como se muestra en la tercer (los arcos a los costados de la “F” están más arriba y son mas grandes en la consulta) y cuarta imagen (la “J” rotada atraviesa a la “A” más arriba en la consulta, y además es más pequeña).

Tabla 2. Consultas Fallidas

Imagen en BD				
Consulta				

El problema de la proporción se podría resolver deformando las imágenes para que ocupen todo el espacio disponible. El segundo problema es más complejo ya que habría que segmentar la imagen primero y luego comparar sus componentes constituyentes y sus posiciones relativas.

Como resultados importantes de este estudio podemos enumerar las siguientes:

- Ante casos reales de 1-Shot Learning, si no se cuenta con una base de datos para realizar transfer-learning con fine-tuning, se pueden utilizar mecanismos de aumentación enriquecidos. Como desventaja, se hace notar que este procedimiento es costoso en tiempo y espacio.
- El uso del algoritmo propuesto de “Relieve” para imágenes de líneas en blanco y negro, permite que la Red Siamesa generalice mucho mejor, alcanzando tasas de acierto relativamente buenas, y que aún pueden mejorarse más.
- Aún así, cuando se agregan nuevos elementos a la base de datos sin reentrenar el modelo, la tasa de aciertos sigue siendo baja como para su implementación en un sistema real.

7. Conclusiones y Trabajo Futuro

En este artículo se presentó el uso de Redes Siamesas y Aumentación como método de extracción de características para la búsqueda de imágenes de marcas de ganado y se propuso una técnica de transformación de las imágenes aumentadas llamada Relieve, que produce mejoras significativas en la extracción de características orientadas a las consultas por similitud. El Relieve consiste en expandir las líneas de las imágenes hacia ambos lados, utilizando valores decrecientes que indican la importancia de cada pixel en la imagen. La aumentación enriquecida con esta nueva técnica mejora notablemente la tasa de acierto de las búsquedas por similitud.

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

- Utilizar técnicas de segmentación para dividir imágenes compuestas en una sola y utilizarlas como subconsultas para mejorar los aciertos.
- Experimentar con otras bases de datos de imágenes de líneas en escala de grises/blanco y negro.
- Modificar la Red Siamesa para que utilice Triplet Loss como función de pérdida.
- Analizar estrategias para que la Red generalice mejor, de tal manera de que sea robusta antes nuevas imágenes sin reentrenamiento.
- Experimentar utilizando la Red Siamesa como Función de Distancia (Metric Learning).
- Desarrollar una aplicación utilizando este método para resolver el problema de búsqueda de similitud de marcas de ganado para el Registro Ganadero del Ministerio de Desarrollo Agrario.

Referencias

- [1] Chávez, Edgar, et al. Searching in metric spaces. *ACM computing surveys (CSUR)* 33.3 (2001): 273-321.
- [2] Decreto Ley Nacional 22939 SENASA (1983). 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 (1999): 56-63.
- [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 (2000): 1349-1380.
- [5] Valova, Irena, Boris Rachev, and Michael Vassilakopoulos. Optimization of the algorithm for image retrieval by color features. *International Conference on Computer Systems and Technologies-CompSysTech.* (2006).
- [6] Sarfraz, Muhammad, and Ahmad Ridha. Content-based image retrieval using multiple shape descriptors. *2007 IEEE/ACS International Conference on Computer Systems and Applications.* IEEE, (2007).

- [7] Pass, Greg, and Ramin Zabih. 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 (2004): 1-19.
- [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. *International Conference on Information Technology: Coding and Computing (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. *IEEE International Conference on Multimedia and Expo, 2001. 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 (2008): 227-237.
- [14] Fukushima, Kunihiko. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks* 1.2 (1988): 119-130.
- [15] LeCun, Yann, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86.11 (1998): 2278-2324.
- [16] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation* 18.7 (2006): 1527-1554.
- [17] Nair, Vinod, and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. *Icml*. (2010).
- [18] Krizhevsky, Alex, Sutskever, Ilya and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)*. Curran Associates Inc., Red Hook, NY, USA, (2012): 1097-1105.
- [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] Springenberg, Jost Tobias, et al. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806* (2014).
- [22] Huang, Gao, et al. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017).
- [23] Larsson, Gustav, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648* (2016).
- [24] Szegedy, Christian, et al. Going Deeper with Convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2015).
- [25] Szegedy, Christian, et al. Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. *Thirty-first AAAI conference on artificial intelligence*. (2017).
- [26] Szegedy, Christian, et al. Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016).
- [27] He, Kaiming, et al. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016).
- [28] Fierro, Atoany N., et al. Redes Convolucionales Siamesas y Tripletas para la Recuperación de Imágenes Similares en Contenido. *Información tecnológica* 30.6 (2019): 243-254.
- [29] Melekchov, Iaroslav, Juho Kannala, and Esa Rahtu. Siamese network features for image matching. *2016 23rd international conference on pattern recognition (ICPR)*. IEEE, (2016).
- [30] Dong, Xingping, and Jianbing Shen. Triplet loss in siamese network for object tracking. *Proceedings of the European conference on computer vision (ECCV)*. (2018).
- [31] Hoffer, Elad, and Nir Ailon. Deep metric learning using triplet network. *International workshop on similarity-based pattern recognition*. Springer, Cham, (2015).
- [32] Hermans, Alexander, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737* (2017).
- [33] Wang, Yaqing, et al. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)* 53.3 (2020): 1-34.
- [34] 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).
- [35] Pan, Sinno Jialin, and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22.10 (2009): 1345-1359.
- [36] Storkey, Amos. When training and test sets are different: characterizing learning transfer. *Dataset shift in machine learning* 30 (2009): 3-28.
- [37] Pan, Sinno Jialin, and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22.10 (2009): 1345-1359.
- [38] Kolesnikov, Alexander, et al. Big transfer (bit): General visual representation learning. *European conference on computer vision*. Springer, Cham, (2020).
- [39] Yu, Haizi, et al. Learning from One and Only One Shot. *arXiv preprint arXiv:2201.08815* (2022).
- [40] Schultz, Matthew, and Thorsten Joachims. Learning a distance metric from relative comparisons. *Advances in neural information processing systems* 16 (2003).
- [41] Bellet, Aurélien, Amaury Habrard, and Marc Sebban. A Survey on Metric Learning for Feature Vectors and Structured Data. *arXiv preprint arXiv:1306.6709* (2013).
- [42] Brisaboa, Nieves R., et al. Similarity search using sparse pivots for efficient multimedia information retrieval. *Eighth IEEE International Symposium on Multimedia (ISM'06)*. IEEE, (2006).
- [43] Aronovich, Lior, and Israel Spiegler. CM-tree: A dynamic clustered index for similarity search in metric databases. *Data & Knowledge Engineering* 63.3 (2007): 919-946.
- [44] Almeida, Jurandy, Ricardo da S. Torres, and Neucimar J. Leite. BP-tree: An efficient index for similarity search in high-dimensional metric spaces. *Proceedings of the 19th ACM international conference on Information and knowledge management*. (2010).
- [45] 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): 721-733.
- [46] Britos, Luis, A. Marcela Printista, and Nora Reyes. DSACL+-tree: A dynamic data structure for similarity search in secondary memory. *International Conference on Similarity Search and Applications*. Springer, Berlin, Heidelberg, (2012).

- [47] Sampallo, Guillermo M., et al. Reconocimiento de marcas de ganado. *IX Congreso Argentino de Ciencias de la Computación*. (2003).
- [48] Sánchez Torres, Germán, and Manuel E. Rodríguez García. Medida de similaridad entre imágenes de marcas de ganado mediante distribuciones de forma. *Revista Ingenierías Universidad de Medellín* 13.25 (2014): 177-189.
- [49] Sanchez, G., and M. Rodriguez. Cattle marks recognition by hu and legendre invariant moments. *ARPN Journal of Engineering and Applied Sciences* 11.1 (2016): 607-614.
- [50] Silva, Carlos, et al. Cattle Brand Recognition using Convolutional Neural Network and Support Vector Machines. *IEEE Latin America Transactions* 15.2 (2017): 310-316.
- [51] Pascal, A., Michel, L., Romani, R. Zernike Moments vs ALR3 Applied to Similarity Searching of Cattle Brands. *VI Seminario Argentina-Brasil de Tecnologías de la Información y la Comunicación - SABTIC (2018)*.