UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

SISTEMA DE RECONOCIMIENTO DE PRODUCTOS BASADO EN IMÁGENES DE GÓNDOLAS DE SUPERMERCADO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN TECNOLOGÍAS DE LA INFORMACIÓN

FRANCISCO JAVIER SÁNCHEZ FUENTES

PROFESOR GUÍA
JOSÉ MANUEL SAAVEDRA RONDO

MIEMBROS DE LA COMISIÓN
NANCY HITSCHFELD KAHLER
MAURICIO CERDA VILLABLANCA
RICARDO BARRIENTOS ROJEL

SANTIAGO DE CHILE
2023

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN TECNOLOGÍAS DE LA INFORMACIÓN
POR: FRANCISCO JAVIER SÁNCHEZ FUENTES
FECHA: 2023
PROF. GUÍA: JOSÉ MANUEL SAAVEDRA RONDO

SISTEMA DE RECONOCIMIENTO DE PRODUCTOS BASADO EN IMÁGENES DE GÓNDOLAS DE SUPERMERCADO

En este trabajo de tesis se aborda la situación que ocurre en una empresa auditora, orientada a hacer estudios en el mercado del retail. La empresa lleva a cabo diversos estudios, tales como la medición de presencia de productos en el retail en general, la disponibilidad de estos en las salas de venta, y su distribución en las góndolas, permitiendo a los fabricantes de productos comprender la oportunidad de ventas y mejorar la gestión de su stock.

Esta empresa tiene a su disposición trabajadores de dos tipos principalmente: supervisores y auditores. El supervisor ayuda a organizar los trayectos de los auditores y a revisar los estudios que serán resueltos por ellos. Por su parte, los auditores son quienes realizan el levantamiento de información de las salas, buscando manualmente los productos solicitados en los estudios.

Realizar un estudio de presencia de manera manual conlleva desafíos, como la dificultad de escanear la totalidad de productos en la góndola, ubicar productos en las góndolas de una categoría específica, y gestionar el tiempo, considerando la lista de productos que un auditor debe abordar durante el estudio.

Este trabajo de tesis busca reducir el tiempo en la captura de información de productos, actividad que se lleva a cabo en los estudios de presencia, mediante la incorporación de un sistema de reconocimiento de productos para góndolas de supermercado. Este sistema permite automatizar la acción de localizar un producto, y reconocerlo mediante una imagen de la góndola del punto de venta. Como resultado, el auditor obtendrá los productos reconocidos y su cantidad en las góndolas.

Para evaluar la solución se realizaron tres experimentos: 1) Reconocimiento de marcas en leches líquidas, con marcas como Colún, Líder y Lonco Leche, y obteniendo un porcentaje de error en la clasificación de 12.5%, 11.8% y 0% respectivamente. 2) Reconocimiento de tipo de producto por marca Colún para los tipos de productos Descremada, Semi descremada y Entera. 3) Flujo completo auditor, donde se analizó una imagen que contiene leches de las marcas Colún, Surlat y Soprole, obteniendo un 75% de acierto sobre los productos de la marca Colún y el total en sus tipos de productos. Como resultado obtenemos una disminución del 34% del tiempo de levantamiento de información.

Las ventajas potenciales que entrega la solución a la empresa auditora, es que ahora no se requiere de un número elevado de muestras etiquetadas para lograr el reconocimiento de un producto, y que es posible contar con un respaldo de los estudios realizados por los auditores, para entregarlos a los fabricantes a modo de informe. Además, permite volver a consultar la información con una referencia en historial de estudios, y volver a reprocesar las imágenes en caso de ser necesario.

Tabla de Contenido

Índice de Figuras

Índice de Tablas

Capítulo 1: Introducción

Este capítulo entrega una idea general de la situación actual de la empresa, el problema abordado y los objetivos del trabajo reportado en este documento de tesis. Además, se incluye una breve descripción de la solución desarrollada, así como sus alcances y limitaciones.

1.1. Contexto del trabajo

La industria del *retail* se encarga de vender al por menor productos de diversos proveedores/fabricantes. Uno de los grandes actores de esta industria son los supermercados. Estos cuentan con centros de distribución, almacenes o depósitos, donde reciben los productos comprados al por mayor. Luego de recibirlos, los productos son enviados a distintos locales comerciales o sucursales, que constituyen los puntos de venta donde el consumidor final adquiere los productos. El flujo de los productos, para que estén disponibles en las tiendas o puntos de ventas, se puede ver en la Figura 1.

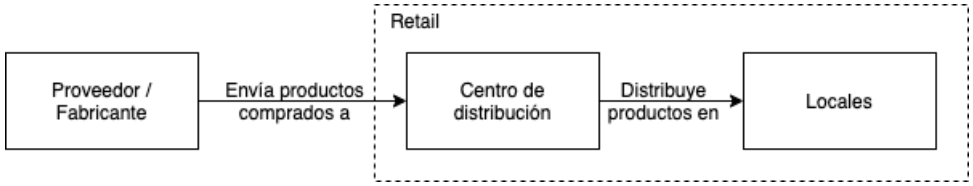

Figura 1: Flujo de los productos de retail.

Cuando se compra un producto nuevo al por mayor, para ser vendido en los supermercados, ocurren dos procesos importantes en la negociación entre los retailers y sus proveedores: 1) establecer el costo del producto (que implica contemplar diversos factores, tales como la demanda, época del año, campañas, descuentos asociados, etc.) y 2) determinar la ubicación del producto dentro del supermercado. A modo de ejemplo, en la Figura 2 se representan los puntos de exhibición de un supermercado. El punto de exhibición más abundante y utilizado son las góndolas, que son las vitrinas más conocidas por los compradores.

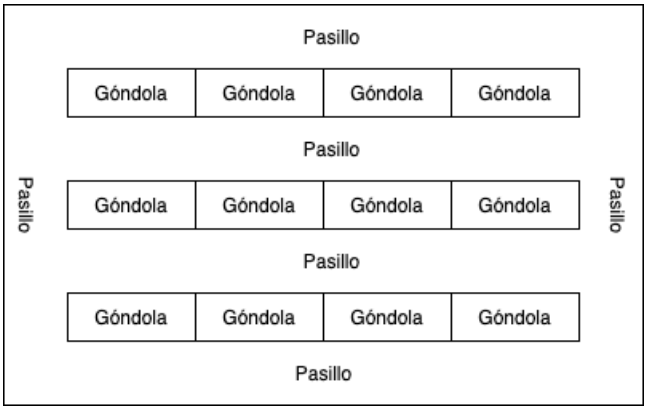

Figura 2: Puntos de exhibición para venta de productos en un supermercado.

Como una forma de verificar que lo negociado se cumpla, los proveedores (fabricantes de los productos) usualmente contratan a empresas auditoras que se encargan de supervisar que sus productos estén en los lugares que fueron acordados. Estas auditorías deben además reportar a los proveedores si su producto no está presente en el punto de venta. A esta actividad le llamaremos *estudios,* y como se mencionó antes, corresponden a procesos de levantamiento de información en los puntos de venta.

Este trabajo de tesis se realiza en el contexto de una empresa auditora de mercado (en adelante simplemente, la empresa, por motivos de confidencialidad), la cual se enfoca en el análisis, control y gestión de puntos de venta. Ésta ofrece principalmente servicios a los proveedores de los retailers, dentro de los cuales se encuentran los servicios de análisis, gestión y control. El servicio de análisis identifica oportunidades de mejora en las ventas, con base en la integración de diferentes fuentes de datos. El servicio de gestión busca priorizar y tomar acción sobre los problemas de los diferentes puntos de venta. Por último, el servicio de control busca elevar los estándares de calidad, mediante la observación y obtención de datos desde el punto de vista del consumidor.

Durante los años 2019 y 2020, la empresa se ha enfrentado al desafío de mejorar sus servicios mediante la transformación digital, logrando consolidar en un único servicio todos los anteriores (análisis, gestión y control). Este servicio integral busca mejorar el valor entregado a sus clientes, a través de la automatización de procesos, realizando la integración del análisis de datos mediante mecanismos de business intelligence (para la creación de reportería) e inteligencia artificial (para la predicción de la venta de los productos).

Actualmente, la empresa inició una segunda etapa de mejora para sus servicios, basado en la automatización de procesos manuales; particularmente, en la obtención de datos en los puntos de venta. Con esto se busca disminuir el tiempo de recolección en el proceso de captura de datos, y reducir el esfuerzo durante la realización de los estudios.

Para realizar un estudio, el auditor debe realizar distintas actividades (usualmente en secuencia): recorrer un punto de venta buscando en las góndolas el producto hasta localizarlo, buscar el código de barra, escanear dicho código para registrar su presencia, responder las preguntas asociadas al estudio en caso de ser requeridas, y cuando termina de procesar los productos de una misma categoría, debe capturar imágenes para respaldar su trabajo. Un auditor puede demorar entre 3 a 4 horas en realizar un estudio considerado como “grande” (1.000 productos aproximadamente).

1.2. Problema abordado

Lo que buscamos abordar en este trabajo de tesis es reducir el tiempo total que ocupa actualmente el auditor para realizar un estudio, el cual es determinado usualmente por la cantidad de productos a procesar. Este tiempo está a su vez relacionado con el proceso de levantamiento de la información, por lo que representa una limitación para que un auditor pueda realizar más estudios durante su jornada laboral.

Tal como se indicó antes, la empresa auditora cuenta con una variedad de estudios que ofrece a sus clientes. Sin embargo, en el marco de esta tesis se aborda únicamente el estudio de presencia de un

producto, el cual involucra el levantamiento de información por producto, como resultado se entrega al cliente un reporte que indica si los productos están quebrados (sin stock en la góndola) o están disponibles para su compra. En el proceso actual (legado), el flujo que debe realizar el auditor para levantar esta información en un estudio de presencia se muestra en la Figura 3.

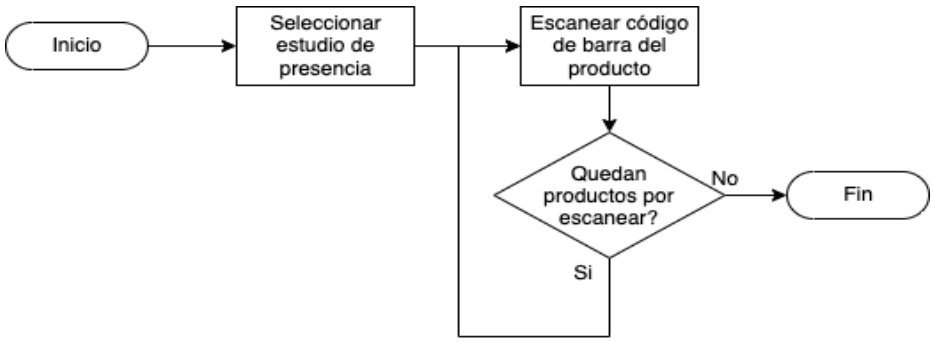

Figura 3: Flujo del proceso actual realizado por el auditor del estudio de presencia (As-Is).

El flujo de la Figura 3 es secuencial por cada producto incluido en la lista de productos a relevar. Un estudio de presencia sobre una lista de 200 productos puede demorar 70 minutos aproximadamente. El problema que se resolverá en esta tesis será apoyar y automatizar parte del proceso de levantamiento de información para varios productos.

1.3. Objetivos de la tesis

El objetivo principal es reducir el tiempo en la captura de información de productos durante los estudios de presencia, mediante la incorporación de un sistema de reconocimiento de productos para góndolas de supermercado. Para alcanzar el objetivo general definido, se debe completar los siguientes objetivos específicos:

- Identificar los productos detectados en una imagen de la góndola de supermercado.
- Implementar un sistema de reconocimiento de productos para la extracción de información desde imágenes capturadas en terreno.

1.4. Resumen de la solución desarrollada

La solución desarrollada cumple el propósito de apoyar el proceso de levantamiento de información, mediante el reconocimiento de productos utilizando modelos de *inteligencia artificial* y *machine learning*, usando fotos de góndolas de supermercado como información base. El sistema de reconocimiento de productos permite automatizar la acción de localizar un producto y reconocer su marca, además de su tipo. Dicho sistema también complementa los resultados, indicando la cantidad de coincidencias encontradas.

Este sistema interviene en el proceso descrito en la Figura 3, donde el *auditor* debe realizar la localización manual del producto, y un escaneo de código de barra con el celular (acción manual de localizar y tomar el producto). Con la solución implementada, la cual se muestra en la Figura 4,

el *auditor* tendrá la opción de realizar una búsqueda automatizada, capturando una foto de la góndola de supermercado. Con el resultado de esta búsqueda se podrá ajustar, en caso de ser necesario, las predicciones realizadas por los modelos. Ahora el auditor tendrá un respaldo de su trabajo, y la empresa auditora podrá volver a reconstruir los estudios desde las imágenes, inclusive pudiendo extraer más información.

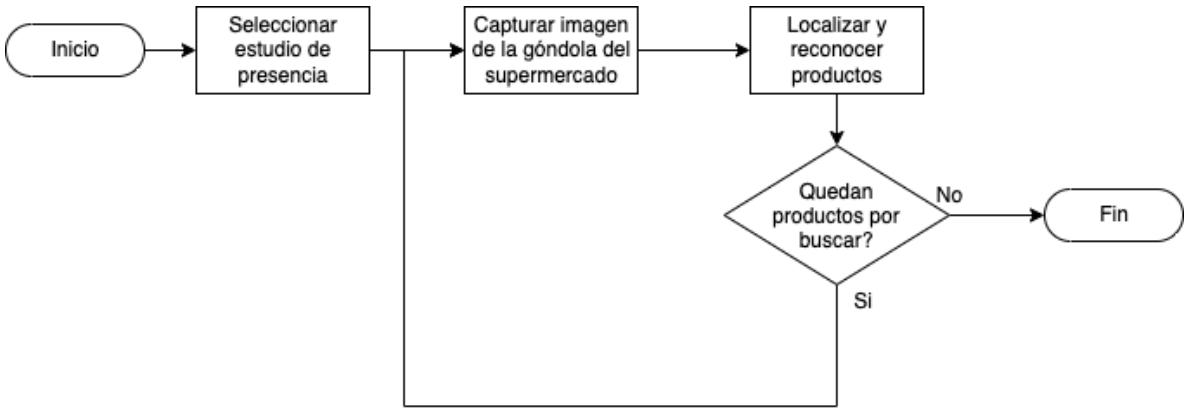

Figura 4: Flujo realizado por el auditor con el sistema de reconocimiento de productos (To-Be).

Para la detección de objetos en las imágenes de las góndolas, se propuso el sistema de detección de productos descrito en el capítulo 3 (particularmente ver la sección 3.3.2: *Diagrama de contenedores del sistema)*. Se consideran dos actores, el auditor que es quien captura imágenes de la góndola de supermercado y realiza los estudios, y el supervisor que es quien se encarga de gestionar y validar los estudios realizados.

El sistema de reconocimiento de productos se basa en la técnica de "pocas muestras" para el renacimiento, complementado con aprendizaje auto-supervisado (referirse al Capítulo 2, sección 2.3.2 *Bootstrap Your Own Latent (BYOL)*), por lo que fue necesario contar con una base de datos de referencia, que incluyera productos validados con sus respectivas etiquetas.

La solución entrega valor a los *fabricantes* brindando visibilidad sobre los estudios y sus resultados. En cuanto a la empresa, ésta se beneficia de automatizar procesos manuales y tener la alternativa de volver a consultar información desde las imágenes capturadas durante los estudios.

Para acotar el alcance de la tesis, el sistema de reconocimiento de productos implementado utilizó imágenes de góndolas de supermercado que corresponden a las categorías de leches. Dada la naturaleza del problema, se escogió a esta categoría dado que es conocida por tener una alta variabilidad de productos dentro de la misma clase.

1.5. Estructura del documento

El presente documento está estructurado en 6 capítulos. El capítulo 2 corresponde al Marco teórico, en éste se describe la situación actual y se presentan trabajos relacionados como alternativas de solución. En el capítulo 3 se muestra la concepción de la solución, junto con los requisitos y el diseño de la solución planteada. El capítulo 4 describe la implementación de la solución, donde se

muestra las funcionalidades del sistema según el perfil de usuario considerado. En el capítulo 5 se muestra la evaluación de la solución, donde se determina qué tan bien la solución permite alcanzar los objetivos definidos. Finalmente, en el capítulo 6 se presentan las conclusiones y el trabajo futuro.

Capítulo 2: Marco teórico

Tal como se mencionó antes, en este capítulo abordaremos en detalle la descripción de la situación actual de la empresa, y a su vez, los conceptos de negocio que son necesarios para comprender el problema, los desafíos técnicos, antecedentes generales, implementaciones existentes, trabajos relacionados y alternativas de solución.

2.1. Descripción de la situación actual

La situación actual se desarrolla en una empresa auditora orientada al mercado de los retails, la cual tiene a su disposición trabajadores de dos tipos principalmente: supervisores y auditores. El supervisor ayuda a organizar los trayectos de los auditores y a revisar los estudios que serán resueltos por ellos. Por su parte, los auditores son quienes realizan el levantamiento de información de las salas, buscando manualmente los productos solicitados en los estudios que se les encomiendan.

Para realizar un estudio, el auditor debe realizar distintas actividades (usualmente en secuencia): recorrer un punto de venta buscando en las góndolas el producto hasta localizarlo, buscar el código de barra, escanear dicho código para registrar su presencia, responder las preguntas asociadas al estudio en caso de ser requeridas, y cuando termina de procesar los productos de una misma categoría, debe capturar imágenes para respaldar su trabajo. Un auditor puede demorar entre 3 a 4 horas en realizar un estudio considerado como "grande" (1.000 productos aproximadamente).

2.1.1. Conceptos de negocio

Para comprender a cabalidad este trabajo de tesis, es necesario conocer los conceptos que se presentan a continuación.
Estudio. Corresponde a un conjunto de productos que son solicitados para levantar información por un fabricante, el cual tiene la necesidad de conocer el estado actual de sus productos en el retail.

Sala. Corresponde al establecimiento en el cual se le realizará el estudio, conocido como retail o punto de venta.

Visita. Corresponde a la acción del auditor de ingresar a una sala y realizar un estudio.

Fabricante/cliente. Se entiende como el origen de la solicitud para realizar un estudio, comúnmente corresponde al fabricante del producto y su interés de conocer el estado de los productos en las salas.

Categoría. Clasificación de producto que es determinada por su propio tipo.

Tipo de producto. Especificación de un producto por unidad de medida o característica.

Producto. Este es el objeto de estudio asociado a una marca que pertenece a un fabricante.

Góndola. Lugar físico donde se encuentran los productos en una sala.

2.1.2. Desafíos técnicos

Como principal desafío se encuentra el reconocimiento de un producto en una góndola de supermercado, del cual se pueden especificar desafíos complementarios, tales como la detección de los productos y la correcta clasificación, además del desarrollo de un sistema que pueda proporcionar un mecanismo automatizado para la realización de los estudios, que permita la gestión y almacenamiento de datos y resultados.

2.2. Antecedentes generales

La detección de un objeto utilizando la mínima cantidad de muestras para su reconocimiento es un desafío aún vigente en el área de visión por computadora, podemos encontrar palabras clave como *One-Shot Detection* (detección de una sola captura) o *Few-shot Detection* (detección de pocas capturas) que agrupan propuestas de soluciones realizadas por otros autores, las cuales están orientadas a este problema, y que se hace mención en este mismo capitulo, en la sección 2.3 de *Trabajos relacionados*. Para comprender el desafío que abordan las propuestas utilizadas en la tesis, debemos también comprender la evolución de los procesos de automatización sobre imágenes que trata de resolver la visión por computadora, y la creciente corriente de *inteligencia artificial* que actualmente trabajan en conjunto para co-crear modelos de clasificación de imágenes y extracción de características, entre otras varias técnicas que permiten la solución de problemas de procesamiento de imágenes.

2.2.1. Visión por computadora

Para una persona detectar un objeto y reconocerlo es una tarea trivial, comprendemos lo que vemos sin un análisis profundo de su forma, brillo, colores o relieves. Sin embargo, para una computadora identificar un objeto significa realizar un análisis de variables que resulta complejo de entender.

En la última década, con el desarrollo de las redes convolucionales, el análisis de un producto, así como la extracción de características relevantes para determinar una clasificación, ha sido delegado a modelos de *inteligencia artificial*. Estos son capaces de entender mediante prueba y error que característica corresponde a una clasificación, dada por los conjuntos de datos utilizados para aprender.

2.2.2. Redes convolucionales

Las *redes neuronales convolucionales* son sistemas computacionales orientados a resolver problemas en el campo de visión por computadora. Éstas son una subclase de una familia de redes neuronales y se caracterizan por tener un buen desempeño en la *extracción de características*. Esto corresponde al proceso de *caracterizar* una imagen, obteniendo una representación de la "observación", y como resultado se obtiene un vector características. En la figura 5 se puede ver una representación de la arquitectura de la red neuronal.

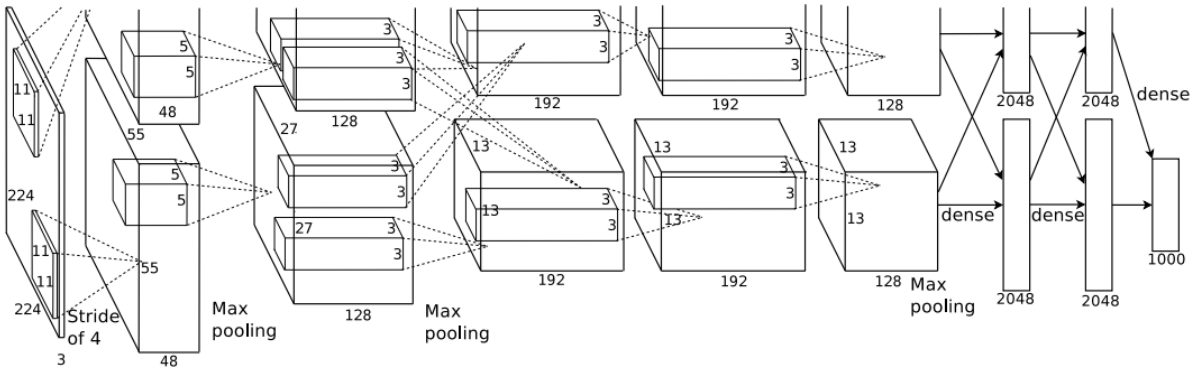

Figura 5: Representación de la arquitectura de la red neuronal convolucional alexnet.

Estas subclases de redes neuronales comenzaron a ser populares en el año 2012, cuando el investigador de la Universidad de Toronto, Geoffrey Hinton y Alex Krizhevsky, participaron en la competencia de *ImageNet*, donde además de ganar la competencia con *Alexnet* [2], también logró reducir el error de forma significativa en el proceso de clasificación. Actualmente Alexnet es considerada la red neuronal convolucional con más influencia en el campo, por su aporte conceptual y práctico. A continuación, se puede ver una tabla comparativa en la Figura 6 del error de clasificación en los próximos años de competencia, también se muestra la reducción significativa del error de clasificación. En el año 2015 el error fue inferior al que comete un humano durante el proceso de clasificación.

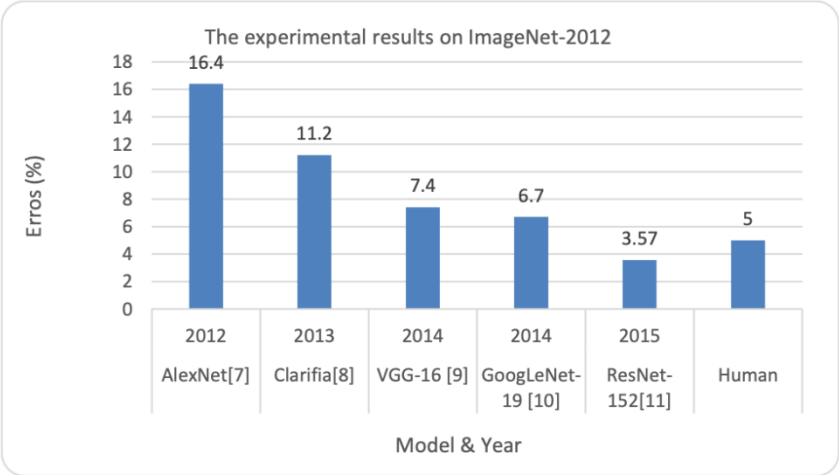

Figura 6: Tabla comparativa de error de clasificación de los modelos ganadores en la competencia de Imagenet (Fuente: [3]).

El esfuerzo realizado en bajar el error de clasificación, fue una hazaña importante para las redes convolucionales en el desafío de *ImageNet.* Esto generó mayor interés en volver a investigar y explorar en nuevos subcampos relacionados con redes neuronales, tales como *clasificación de imágenes*, *detección de objetos*, *reconocimiento de objetos*, y *segmentación semántica*, entre otros.

2.2.3. Extracción de características

La extracción de características es un paso importante dentro de las redes convolucionales en la recuperación, el procesamiento y extracción de datos de imágenes. Este es el proceso de extraer información relevante desde una imagen, buscando reflejar el contenido intrínseco de un dato o conjunto de datos lo más completo posible, para luego ser utilizado por la computadora en la clasificación de una imagen.

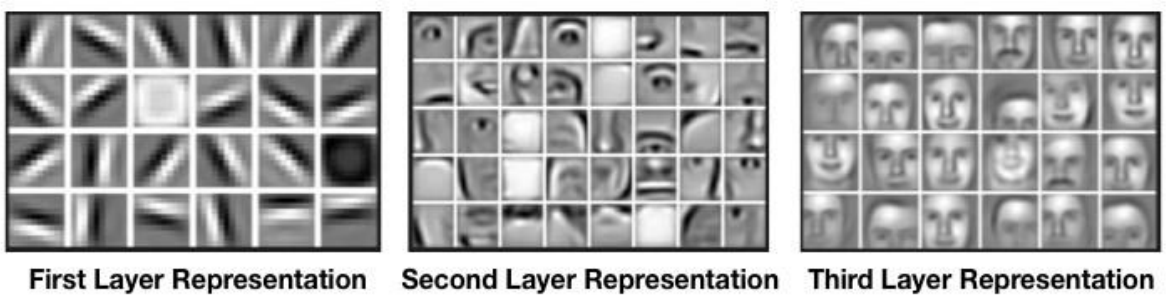

Figura 7: Representación de extracción de características de una red neuronal profunda.

En la Figura 7 se muestra cómo las capas internas de una red convolucional captan datos relevantes para la clasificación de rostros. Ahí se puede apreciar cómo la primera capa (First Layer Representation) encuentra figuras geométricas relevantes para clasificar un rostro. En la medida que se hace más profundo el análisis, vemos cómo la primera capa se complementa con la segunda capa (Second Layer Representation), encontrando secciones tales como, los ojos, nariz y boca. Esto permitirá luego generar una representación vectorial para la identificación de un rostro. A medida que profundizamos en la extracción, podemos ver que la tercera capa (Third Layer Representation), que obtiene los contornos para complementar una representación generalizada del rostro.

2.2.4. Detección de objetos

La detección de objetos es el paso previo a la clasificación del mismo. Inicialmente el área de visión por computadora tuvo como primer desafío comprender un objeto y clasificarlo en el resultado esperado por el algoritmo, como se muestra en la Figura 8. Ahora la detección de objetos busca detectar todas las instancias de las clases predefinidas con la que entrenamos un modelo, y proporcionar su localización aproximada, como se muestra en la Figura 9.

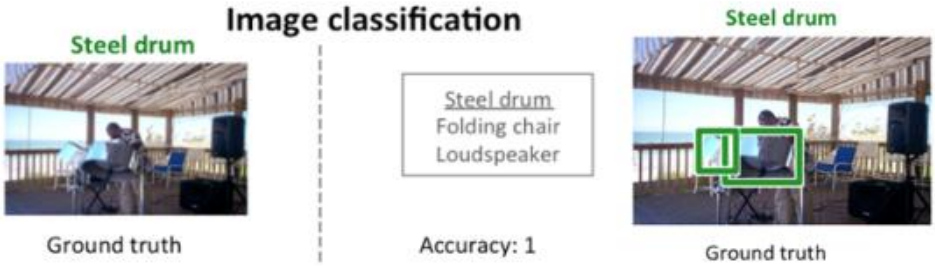

Figura 8: Comparación entre clasificar una imagen según un objeto (lado izquierdo), versus localizar un objeto en la imagen (lado derecho)

Para hacer uso de la detección de objetos, se utiliza un sistema de cajas que encierran al objeto localizado. Es importante comprender que no se requiere de una clasificación para obtener una localización, en otras palabras, encontrar el objeto no significa que conozcamos su clase. La manera de señalar la ubicación de un objeto en una imagen se conoce como *bounding boxing*. Esto se muestra en la Figura 9, para indicar la ubicación guiada por las coordenadas de un plano cartesiano de los objetos encontrados en la imagen.

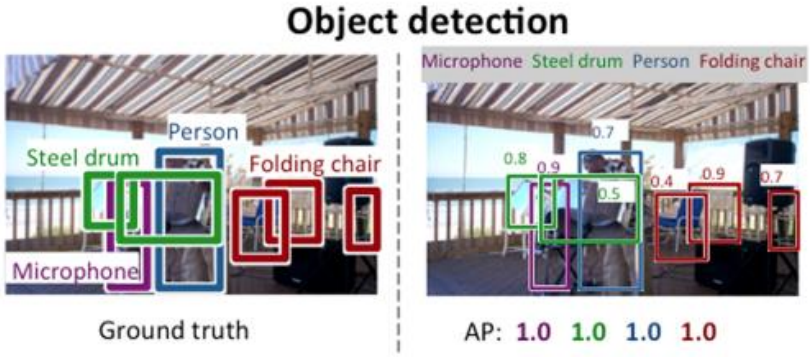

Figura 9: Ejemplo de detección de objetos en una imagen

Bounding Box es un rectángulo abstracto que actuará como punto de referencia visual para la detección de un objeto. Éste representa el candidato con menor error en la selección de su clase.

2.2.5. Clasificación de imágenes auto-supervisada por agrupación (clúster)

La clasificación de imágenes a gran escala, es un desafío complejo que aún no tiene una solución pública o comercial estable y al mismo tiempo escalable. Los costos de etiquetar muestras son altos, y aumentan a medida que la variabilidad de las clases aumenta o si se considera la intra-variabilidad de una misma clase como puede ser en productos con distintos envases. Además, la clasificación *auto-supervisada* de un conjunto de imágenes representa un desafío aún mayor, con rendimientos mucho más débiles en comparación con los modelos supervisados, donde se puede "guiar" a la respuesta enseñando etiquetas específicas.

Como referencia podemos considerar que la precisión de un modelo auto-supervisado es de un 39% con el conjunto de datos de ImageNet, y si utilizamos alguna técnica de clusterización, entonces se logra un 46% para el mismo conjunto de clases [8].

En el caso de utilizar modelos auto-supervisado, las técnicas de agrupación permiten complementar la extracción de características de las redes convolucionales, identificando grupos específicos de patrones según su similitud, usando la distancia euclidiana (como se muestra en la Figura 10), y agrupándolos para ser identificados por una etiqueta posteriormente declarada.

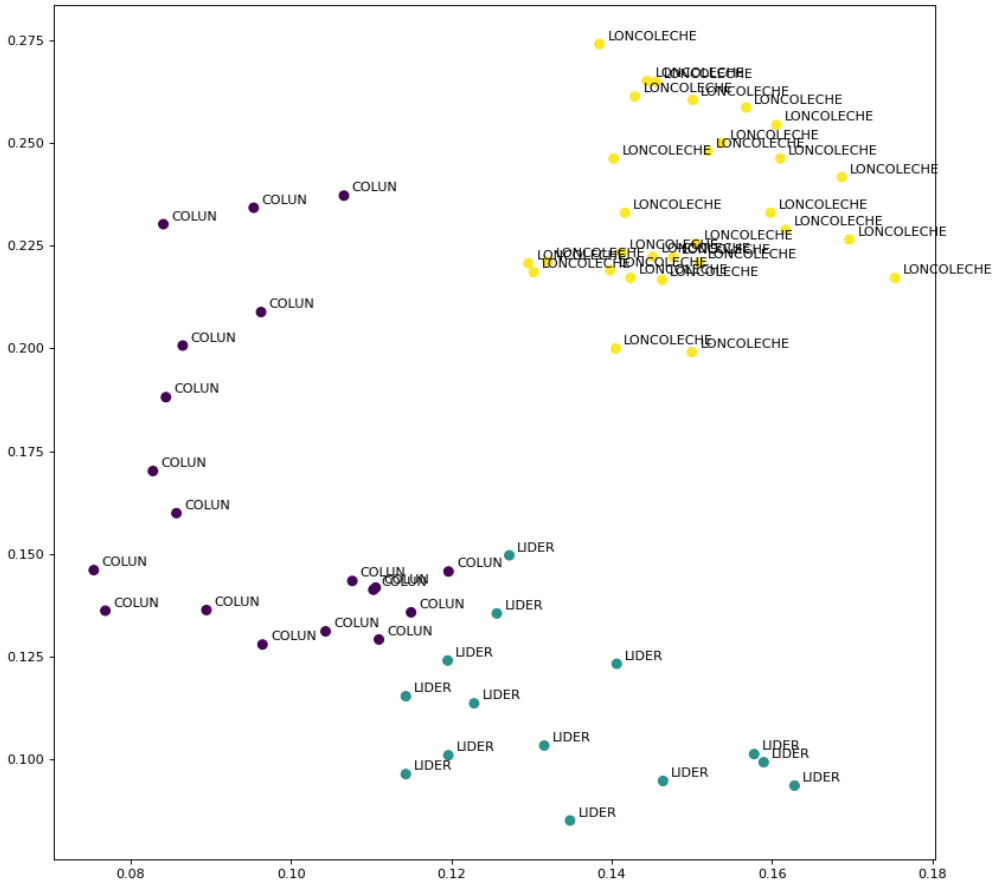

Figura 10: Grupo de imágenes de productos clasificados por la marca del producto utilizando una red convolucional para extraer las características y agruparlas por su similitud.

En la clasificación auto-supervisada de los vectores que contienen la extracción de características se suelen utilizar modelos de clusterización como k-means. Éste utiliza la distancia euclidiana (ecuación (1)), donde se calcula la distancia entre el vector de características y los centroides declarados aleatoriamente en el plano cartesiano. Los centroides son centros de coordenadas x e y, que representarán un punto en común entre un conjunto de vectores de imágenes.

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{1}$$

Para comprender la manera cómo funciona el algoritmo de *k-means*, debemos considerar que es iterativo, por lo que generará una cantidad especificada por parámetros de *centroides*, que serán puestos en un plano cartesiano aleatoriamente. Entonces, por cada iteración se realizará (en cada *centroide)*, el cálculo de las distancias utilizando la ecuación (2) sobre los puntos más cercanos. Se repetirá el proceso hasta que los *centroides* no se trasladen más, o el total de iteraciones definidas se complete.

$$\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \tag{2}$$

Al finalizar el cálculo de los *centroides* obtenemos coordenadas que serán candidatas a un grupo de imágenes, que deben ser representadas por una etiqueta determinada (proceso manual). De esta forma se obtiene la clasificación auto-supervisada de un grupo de imágenes.

2.3. Trabajos relacionados

El problema que aborda la presente tesis no es un área de investigación nueva, a medida que progresan las arquitecturas de los modelos de *inteligencia artificial* principalmente, también se desarrollan innovadoras propuestas que se aproximan a resolver la detección de productos en supermercados o problemas similares sobre detección en imágenes densas (bastantes objetos contiguos), e incluso detección con pocas muestras en un grupo extenso de imágenes con clases variadas.

A continuación, se describen dos modelos existentes que son fundamentales para el desarrollo de la presente tesis: Precise Detection in Densely Packed Scenes y Bootstrap Your Own Latent. Luego se presenta una solución comercial (Amazon Rekognition).

2.3.1. Precise Detection in Densely Packed Scenes

La técnica llamada *Precise Detection in Densely Packed Scenes* [11] es utilizada para la detección de objetos en escenas densamente empaquetadas y que contienen numerosos objetos. La técnica consiste en permitir mayor cantidad de detecciones candidatas sobre un mismo producto, luego se realiza una elección sobre el solapamiento de los candidatos mediante una operación *gaussiana*, reflejando los puntos de mayor confianza para la detección de los productos. En la Figura 13 se muestran los pasos realizados por el modelo propuesto.

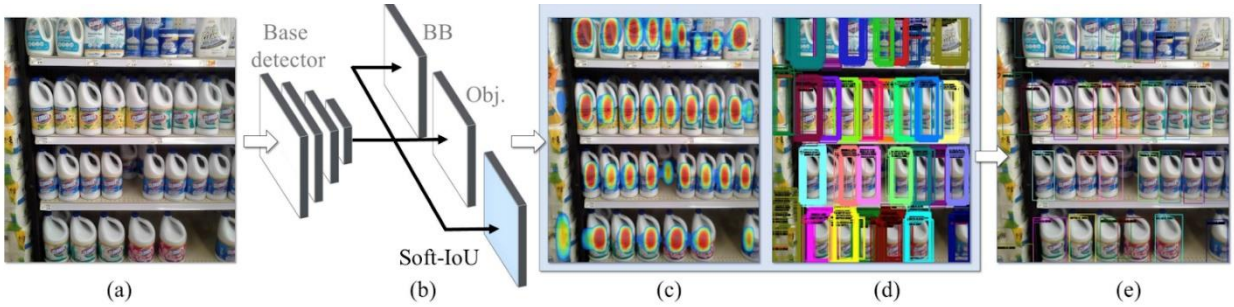

Figura 11: Diagrama del sistema. (a) imagen de entrada. (b) red neuronal base con propuesta de Soft-IoU. (c) EM-Merger convierte Soft-IoU en un mapa de calor gaussiano que representa (d) objetos capturados por varios cuadros delimitadores superpuestos.

El detalle del proceso que realiza la técnica Precise Detection in Densely packed scene es el siguiente.

Soft-IoU score

En la arquitectura de la red neuronal se agrega una tercera capa al final de cada detector. Entonces, dada una cantidad N de detecciones predichas, se obtiene el IoU Score (intersection over unión), el cual es una ponderación del área acertada en la predicción dividido en el total del área cubierta por la predicción de los candidatos (figura 12).

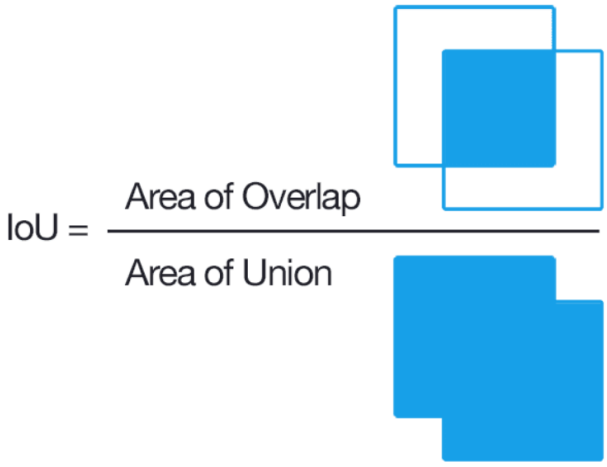

Figura 12: Computing the Intersection over Union

EM-Merged unit

La unidad de inferencia EM-merged utiliza los *bounding box* obtenidos con el *Soft-IoU* que a menudo terminan agrupados sobrepuestos entre sí. La función de la unidad de inferencia EM-Merged es filtrar, fusionar o dividir los grupos de detección sobrepuestos, con el fin de resolver la detección única por objeto.

Figura 13: Detección de productos realizado con el modelo Precise Detection in Densely Packed Scenes

Como se puede muestra en la Figura 13, los resultados sobre una imagen con productos de una góndola de supermercado representan una solución viable. Esta técnica permite la detección sin etiquetas de las localizaciones, considerando la densidad de paquetes presente en la imagen.

2.3.2. Bootstrap Your Own Latent

Bootstrap Your Own Latent (BYOL) [11], es un nuevo enfoque para el aprendizaje de *representación de imágenes* auto-supervisado. En el área del aprendizaje auto-supervisado para imágenes, es un desafío determinar qué técnica utilizar para comprender la representación de una imagen. Existen propuestas como el uso de pares negativos, la cual se muestra en la Figura 14, pero éstas tienen limitaciones ya que son sensibles al contraste de la imagen. Por otro lado, BYOL propone trabajar con vistas aumentadas de la misma imagen, utilizando dos redes convolucionales que se apoyan una a la otra durante el proceso de aprendizaje.

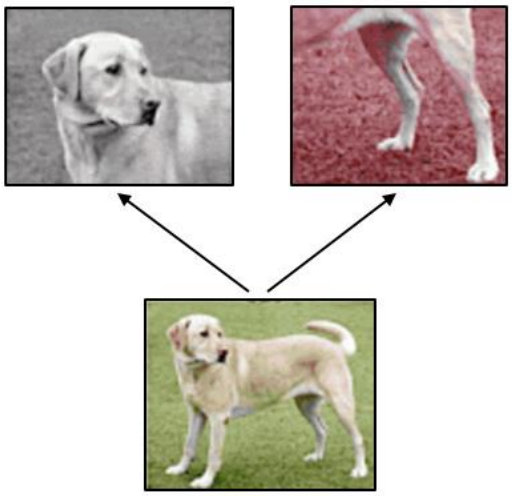

Figura 14: Aprendizaje basado en pares negativos

Dada las dos redes que utiliza BYOL las cuales trabajan en paralelo y dada la predicción encontrada por la red *online*, es comparada con la predicción realizada por *target*, pero solo la red denominada *online* rectifica sus *pesos* para consolidar su aprendizaje, a diferencia de la red conocida como *target* que no considera el feedback y su forma de aprender es "copiando" o "extrayendo" parcialmente el aprendizaje de la red *online*.

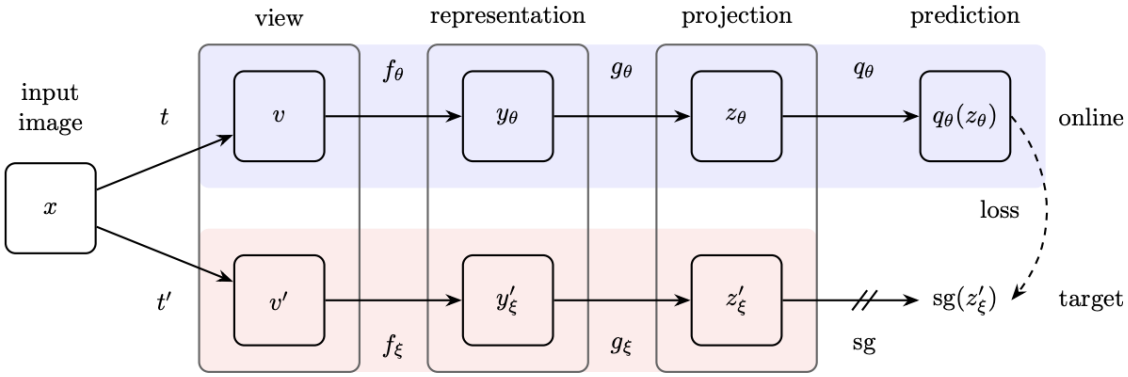

Figura 15: Arquitectura de BYOL.

En la Figura 15 se muestra cómo, desde una imagen (denotado por el símbolo x), se obtienen 2 cortes distintos que serán clasificados por ambas redes en base a unos labels virtuales y se ajustarán los pesos de la red *online*. Esto se hace para reducir la función de error, y posteriormente la red *target* obtendrá parte del conocimiento de la red *online*.

Capítulo 3: Concepción de la solución

En este capítulo abordaremos en detalle el diseño de la solución, y cómo la unión de sus partes permitirá concluir el objetivo de la tesis. En la sección de *requisitos y restricciones de la solución,* veremos qué situaciones debe cumplir el sistema y cuáles son sus restricciones debido a la naturaleza del problema. En la sección de *perfiles de usuario soportados* detallaremos los controles que debe hacer cada perfil según su responsabilidad. En *arquitectura/estructura de la solución* veremos como la unión de las partes de la solución interactúan para obtener resultados, y en la sección *modelos de inteligencia artificial* y *modelos de machine learning* entenderemos la función que cumplen dentro de los procesos de obtención de información.

3.1. Requisitos y restricciones a la solución

Para la concepción del sistema de reconocimiento de productos debemos cumplir los siguientes requisitos:

Requisito Nº *1.* Un conjunto de datos que contenga muestras que representen un producto de góndola de supermercado.

Requisito Nº *2.* Permitir capturar imágenes de góndola de supermercado para su análisis.

Requisito Nº *3.* Permitir crear nuevos productos para ser integrados a los modelos de inteligencia artificial y machine learning.

Requisito Nº *4.* Encontrar los productos e identificarlos en la imagen de la góndola.

Requisito Nº *5.* Devolver la cantidad de productos con su marca y su tipo.

Para la correcta operación del sistema se deben considerar restricciones durante su funcionamiento, como las siguientes:

Restricción Nº *1.* Se debe considerar una carga inicial de imágenes (al menos una), para cada producto con su marca y su tipo que se quiera detectar en el sistema.

Restricción Nº *2.* El enfoque de quien toma la foto debe estar alineado con los extremos de la góndola y no debe estar en movimiento al tomar la captura.

Restricción Nº *3.* El sistema no contempla detectar productos en otro espacio que no sea una góndola de supermercado.

3.2. Perfiles de usuario soportados

A continuación, se indican los perfiles soportados por el sistema de reconocimiento de productos y sus actividades esperadas.

Auditor. Trabajador encargado de realizar estudios solicitados por los clientes. Debe recopilar los datos necesarios para determinar indicadores significativos para el estudio.

Supervisor. Trabajador encargado de supervisar a un grupo de auditores, crear los estudios solicitados por el cliente y asociar las métricas que deben resultar de los datos obtenidos. Además, tiene que realizar soporte en la logística de los mismos auditores durante la realización de sus tareas.

3.3. Arquitectura/estructura de la solución

En la siguiente sección se muestra cómo interactúan los componentes de la solución para realizar el reconocimiento de los productos. En el diagrama de contexto (expresado en C4 [13]), veremos cómo nuestros actores (*auditor* y *supervisor*) usan el sistema, y luego en el diagrama de componentes, veremos en detalle la solución técnica que hay detrás de las solicitudes.

3.3.1. Diagrama de contexto del sistema

Para representar a los actores del sistema y los casos de uso que realizan sobre el sistema de reconocimiento de productos, veremos el siguiente diagrama de contexto que la figura 16.

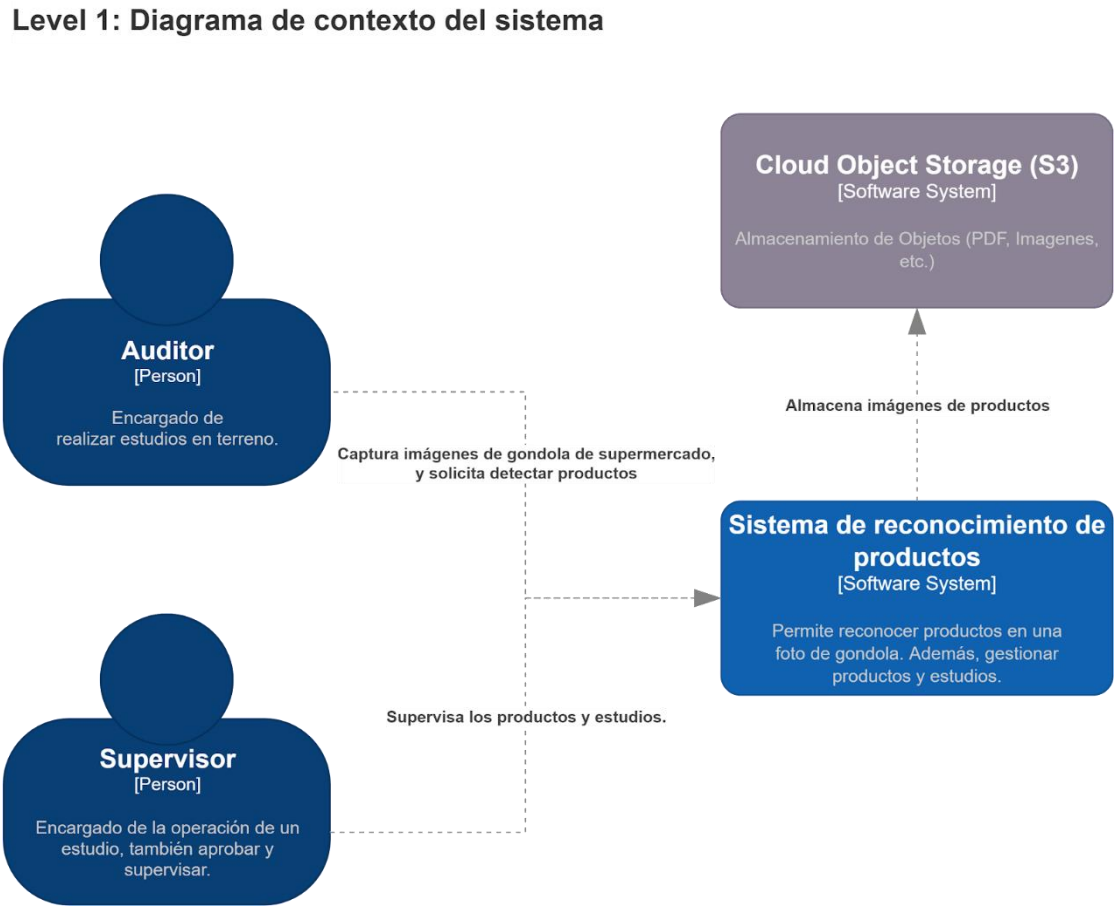

Figura 16: Diagrama de contexto del sistema.

La figura 16 corresponde un diagrama de contexto de un sistema basado en C4 Model [13], donde los actores consultan un sistema de reconocimiento de producto que guarda un registro de las imágenes en un almacenamiento en la nube conocido como Object Storage.

3.3.2. Diagrama de contenedores del sistema

En la Figura 17 se representan los contenedores que serán utilizados por el sistema de reconocimiento de productos para la propuesta de solución.

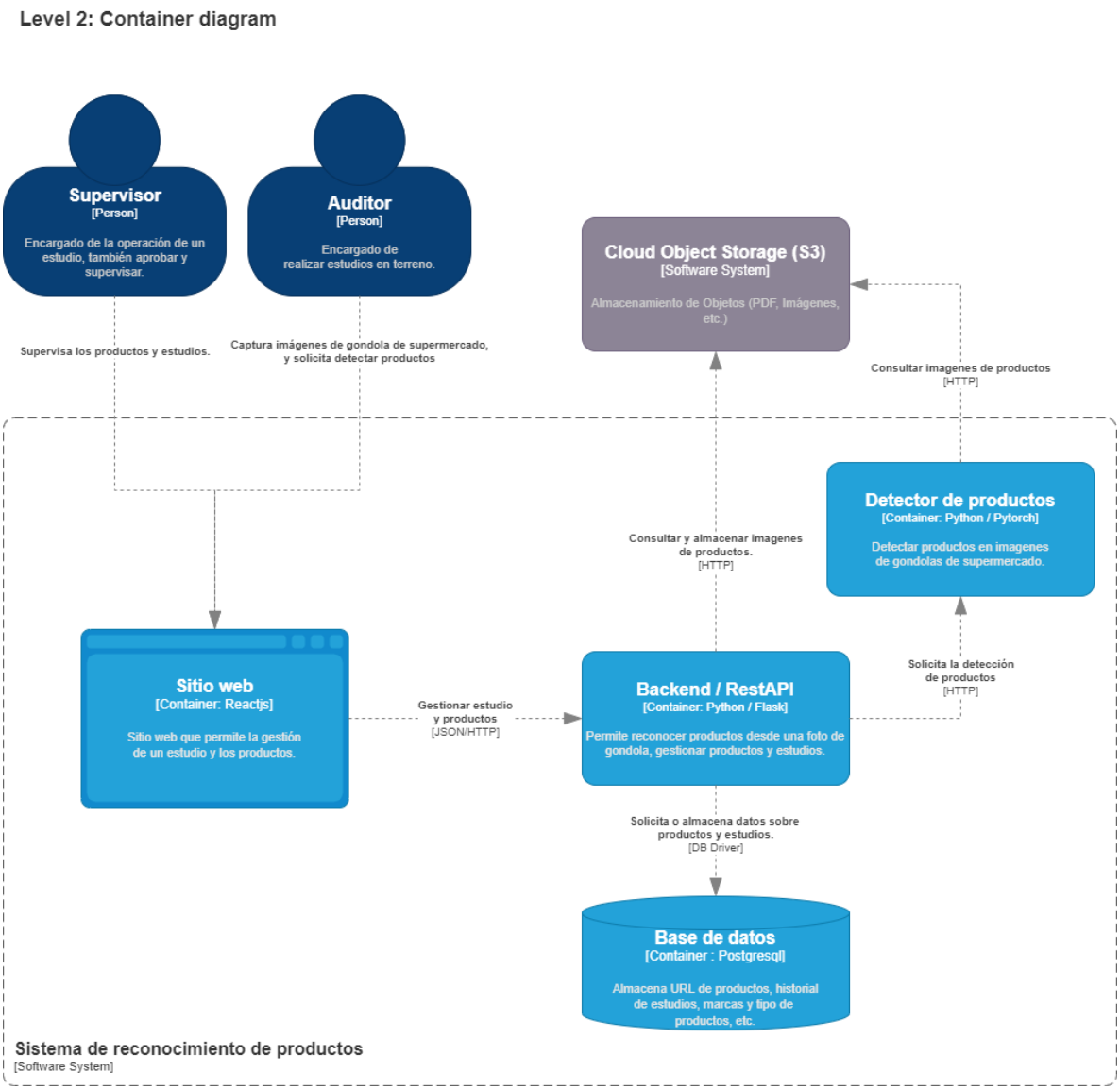

Figura 17: Diagrama de contenedores del sistema.

En la figura 17, se utiliza c4 model [13], donde actores como *supervisor* y *auditor* consultan un sitio web para gestionar el reconocimiento de productos, el cual gestiona mediante un servidor (backend) los procesos necesarios para detectar un producto y utiliza como respaldo un Obejct Storage y una base de datos cuidando la integridad de los datos, además de proveer un histórico de consultas.

3.4. Modelo de datos

El sistema requiere persistir datos para las revisiones posteriores que serán realizadas por los supervisores, también debe almacenar la dirección donde estarán guardadas las imágenes tomadas por los auditores y otros datos de utilidad para el estudio. Para ello se diseñó un modelo de datos (Figura 18) que permite registrar estudios, búsquedas realizadas por el sistema de reconocimiento de productos, perfiles de usuario y productos con sus marcas y sus tipos.

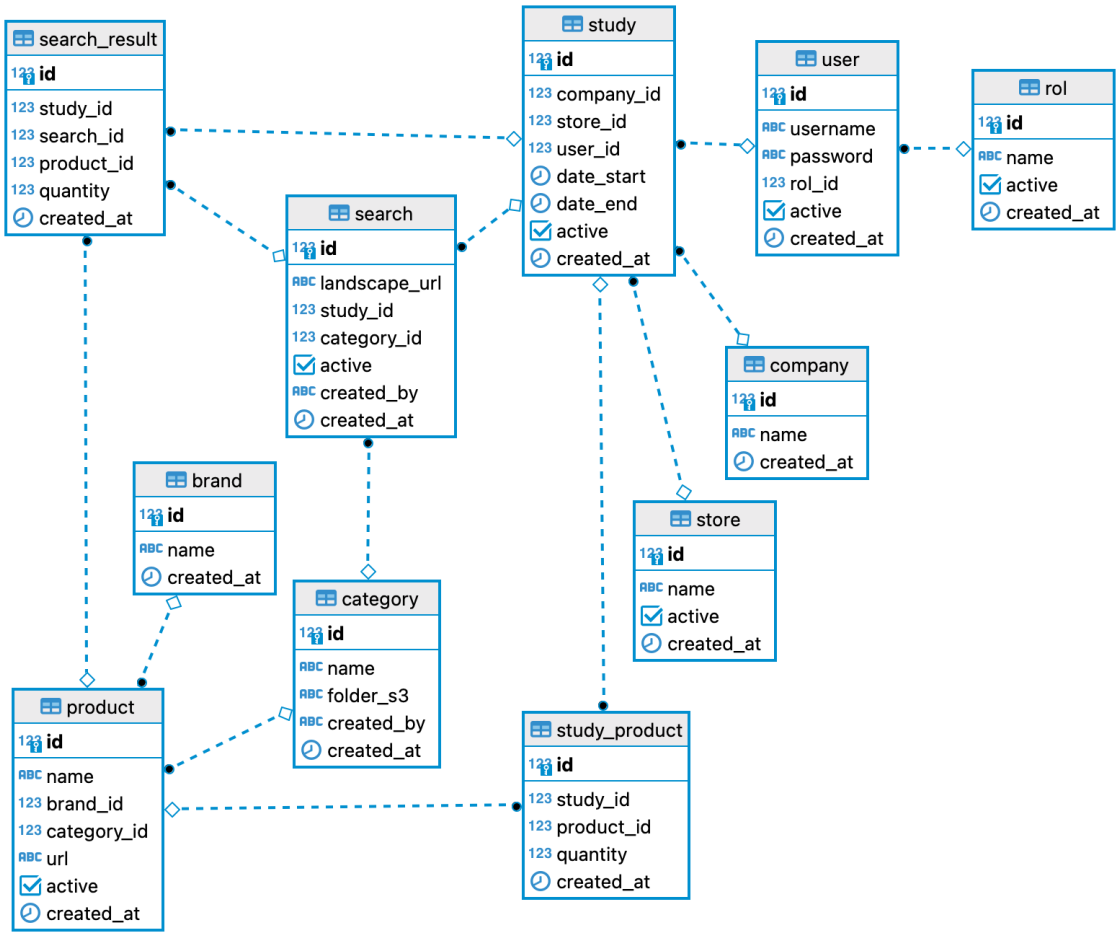

Figura 18: Diagrama de tablas para sistema de reconocimiento de productos.

A su vez, el diseño permite agregar nuevas características (de ser necesario) complementarias a productos, usuarios, estudios, categorías y salas.

3.5. Descripción del conjunto de datos

El conjunto de datos (dataset) utilizado para la solución contempla la categoría de *leches líquidas*, la cual contiene 20 marcas. En la Figura 19 se muestra un gráfico de frecuencia por marca de producto.

En el conjunto de datos se pueden ver particularidades abordadas por el problema, tales como la densidad de productos contiguos o agrupados, con el objetivo de captar la atención de clientes. En la Figura 20 se muestran grupos de imágenes que demuestran densidad de productos en zonas específicas.

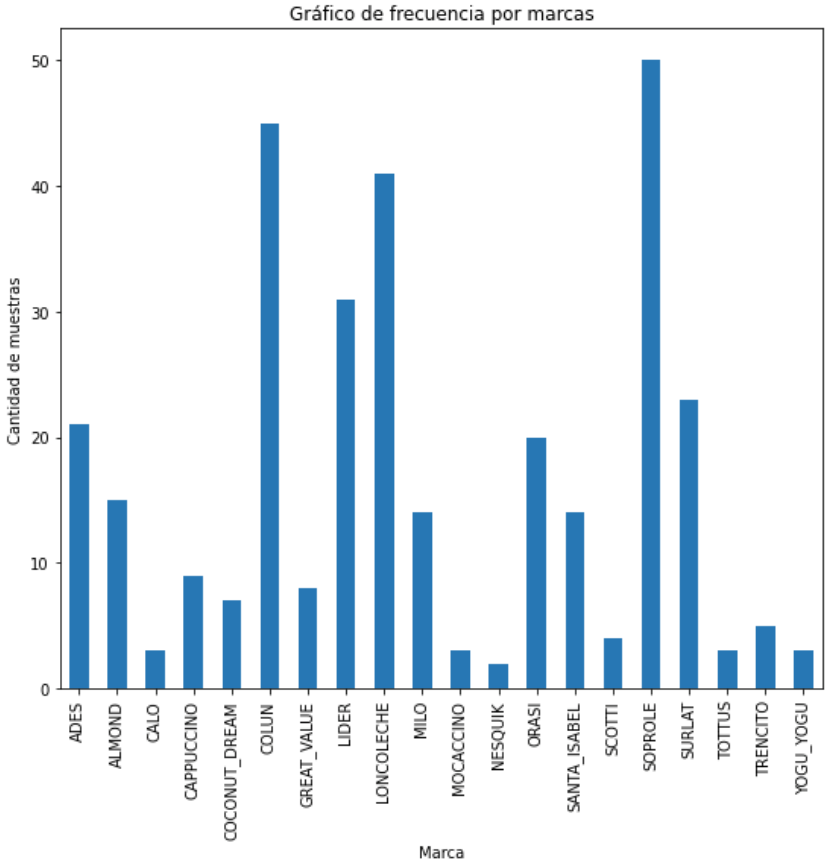

Figura 19: Gráfico de frecuencia por marca de producto

Figura 20: Cajas diseñadas para dar visibilidad a la marca asemejándose al producto en exposición.

La Figura 21 se muestran diseños utilizados para captar la atención del usuario, lo cual dificulta la identificación de los productos.

Figura 21: Densidad de productos en una góndola de supermercado.

Allí se muestra la variabilidad de marcas, diseños y también las opciones de sabores que ofrece un fabricante sobre una línea de producto. Además, se puede ver que la posición de los productos puede no favorecer la detección de estos, particularmente, en el caso de los productos que están en la parte inferior derecha de la góndola (Fig. 21).

3.6. Modelo de inteligencia artificial

En esta sección se detalla a grandes rasgos la funcionalidad que brindan los modelos de *inteligencia artificial* y *machine learning*, y cuán relevante es su impacto en la solución.

3.6.1. Uso de Precise Detection in Densely Packed Scenes

El modelo de *Precise Detection in Densely Packed Scenes*, se encargará de detectar objetos en la góndola, éste realiza la detección de objetos sin conocer marca ni tipo, solo reconoce formas como se muestra en la Figura 22. Como resultado entregará un archivo en formato CSV donde especifica el nombre de la imagen y las posiciones x1, x2, y1 e y2 del objeto. Posteriormente, la información dentro del archivo CSV permitirá obtener individualmente cada producto desde la foto de la góndola de supermercado.

Figura 22: En el lado izquierdo podemos ver una imagen sin procesar, y en el lado derecho se puede ver los resultados del modelo utilizando bounding box.

3.6.2. Uso de Bootstrap your own latent

Una vez obtenido el archivo CSV con el nombre de la imagen y la posición del objecto, se deben obtener individualmente cada objeto (que es un posible producto). Dicho archivo se obtiene de acuerdo a lo descrito en la sección 3.6.1.

Luego, se utilizará el modelo Boostrap your own latent para obtener las características relevantes desde las imágenes de los objetos, estas características nos permitirán entregar información útil para buscar similitudes entre los objetos permitiendo crear grupos según su similitud (por ejemplo, su forma, diseño, letras de la marca, etc.).

Para realizar las agrupaciones se utilizará un modelo de Machine Learning descrito en la sección 3.7.

3.7. Modelo de Machine learning

A diferencia de los modelos de *inteligencia artificial* vistos anteriormente, k-means es un algoritmo iterativo que permitirá el agrupamiento de características similares entre los productos.

Posterior a la extracción de características realizada por BYOL, obtenemos una lista de vectores de características de los productos. Ahora debemos crear un cluster k-means que ubicará los vectores en un plano cartesiano e iterará los centros, calculando las distancias entre los vectores hasta converger, lo cual significa que los centros ya no se moverán más o que llegó al número máximo de 300 iteraciones (valor por defecto).

Con los productos ya agrupados debemos agregar una etiqueta a los centros designados para un grupo de vectores de características según nuestro criterio, para posteriormente utilizarlo como respuesta en la detección de objetos.

3.8. Tecnologías involucradas

Para el desarrollo y puesta en marcha de la solución descrita en la sección *Arquitectura/estructura de la solución,* se requiere de las siguientes tecnologías.

3.8.1. Frontend y Backend

A continuación, se indican las tecnologías utilizadas para implementar el frontend y el backend de la solución.

ReactJS. Esta es una librería de Javascript para desarrollar interfaces de usuario. Se utilizará para implementar la capa de presentación que utilizará el supervisor y el auditor.

Flask. Es un micro web framework escrito en el lenguaje de programación Python. Se utilizará para implementar la capa de lógica del sistema, gestionará los accesos a las bases de datos, la carga de productos y sus imágenes, y la obtención de los productos detectados por el modelo.

PostgreSQL. Motor de base de datos utilizado para almacenar los datos del sistema de reconocimiento de productos.

3.8.2. Modelos de inteligencia artificial / machine learning

A continuación, se listan las tecnologías utilizadas para el desarrollo y prueba de los modelos de inteligencia artificial y machine learning.

Pytorch. Este es un framework de machine learning de código abierto para la creación de prototipos e implementación en producción de soluciones de inteligencia artificial.

Sklearn. Esta es una librería con herramientas simples y eficientes para el análisis predictivo de datos.

JupyterLab. Este es un entorno de desarrollo interactivo basado en la web. Se utilizará para realizar los experimentos sobre los modelos y comprender los resultados obtenidos.

Google Colab. Este es un entorno de desarrollo interactivo que permite combinar código ejecutable y texto enriquecido en un solo documento. Es un servicio que provee la empresa Google y permite utilizar servidores con características útiles para entrenar modelos en menor tiempo.

3.8.3. Infrastructura / PaaS

A continuación, se listan las plataformas y servicios utilizados para realizar los experimentos y pruebas del sistema.

Netlify. Esta es una plataforma como servicio que dispone de un *content delivery network* que permitirá la distribución del sitio web para ser utilizado por auditores y supervisores.

EC2. Elastic compute cloud, es utilizado para desplegar el modelo de inteligencia artificial donde se procesa la captura de imagen realizada por el auditor, también se realizará la detección de los productos.

Capítulo 4: Implementación de la solución

En este capítulo, abordaremos una implementación propia de la solución del Sistema de reconocimiento de productos. Esto corresponde a una aplicación web, que a su vez se puede ver en dispositivos móviles para que pueda ser utilizada por los auditores. El auditor podrá acceder a ver sus estudios, realizar estudios por categoría, captura de imagen de los productos, confirmar información de productos, ver historial y cargar nuevos productos. A continuación, se detallarán las funcionalidades.

4.1. Interfaces

En esta sección se mostrarán las interfaces que incluye el Sistema de reconocimiento de productos a través de screenshots y una breve explicación de las interfaces.

4.1.1. Interfaz de login

En primer lugar, se podrá acceder a la interfaz de login como se muestra en la Figura 23. El usuario que tenga el perfil de auditor, podrá acceder mediante su usuario y contraseña a la aplicación.

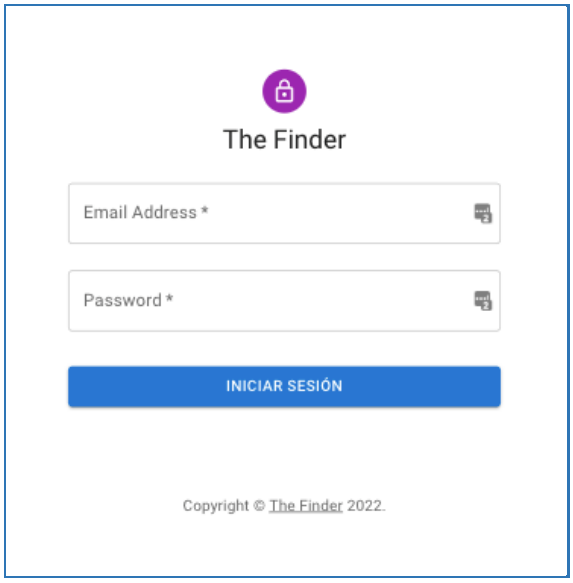

Figura 23: Interfaz de login.

4.1.2. Interfaz mis estudios

Luego de ingresar al sistema con sus credenciales correspondientes, el auditor accede al *Home*, que contiene un listado de los estudios a realizar con la fecha, fabricante que solicita y la sala correspondiente que fueron previamente asignados por su supervisor (ver figura 24).

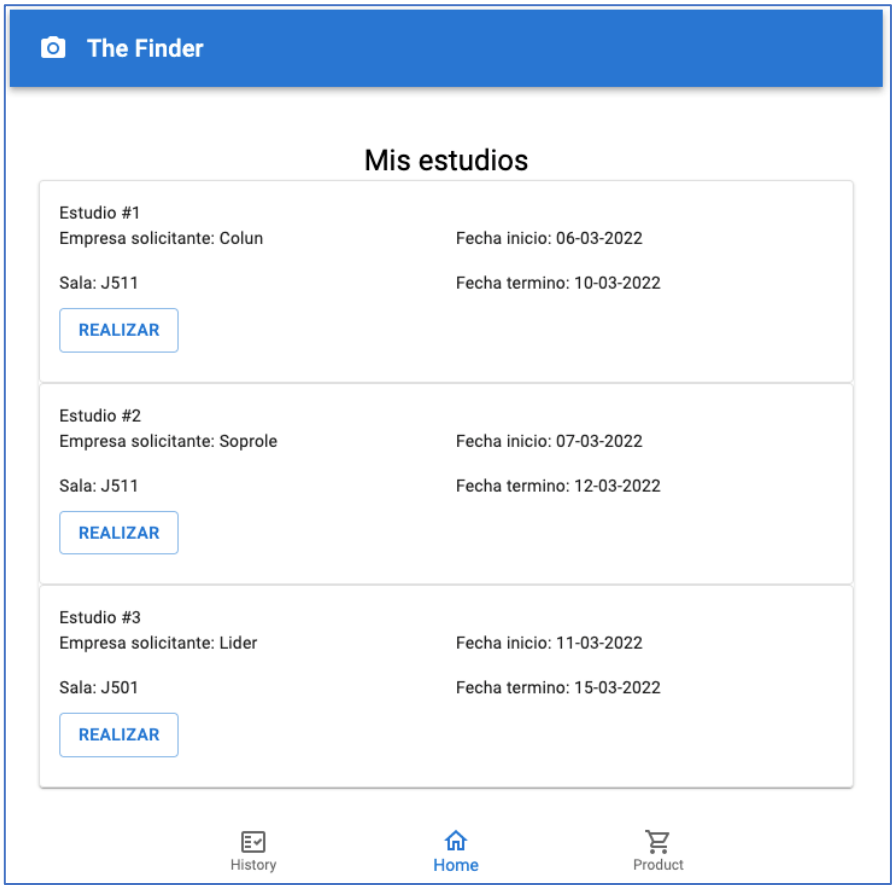

Figura 24: Interfaz mis estudios

El auditor deberá dirigirse a la sala y podrá presionar el botón realizar para comenzar el estudio.

4.1.3. Interfaz realizar estudio por categoría

Una vez seleccionado el estudio, el auditor podrá ver todas las categorías donde debe escanear los productos. Para esto deberá seleccionar la categoría que se encuentre pendiente, las categorías que ya fueron realizadas se mostrarán con un check (ver Figura 25). El auditor deberá dirigirse a la góndola de la categoría seleccionada y podrá presionar el botón “ir” para escanear los productos.

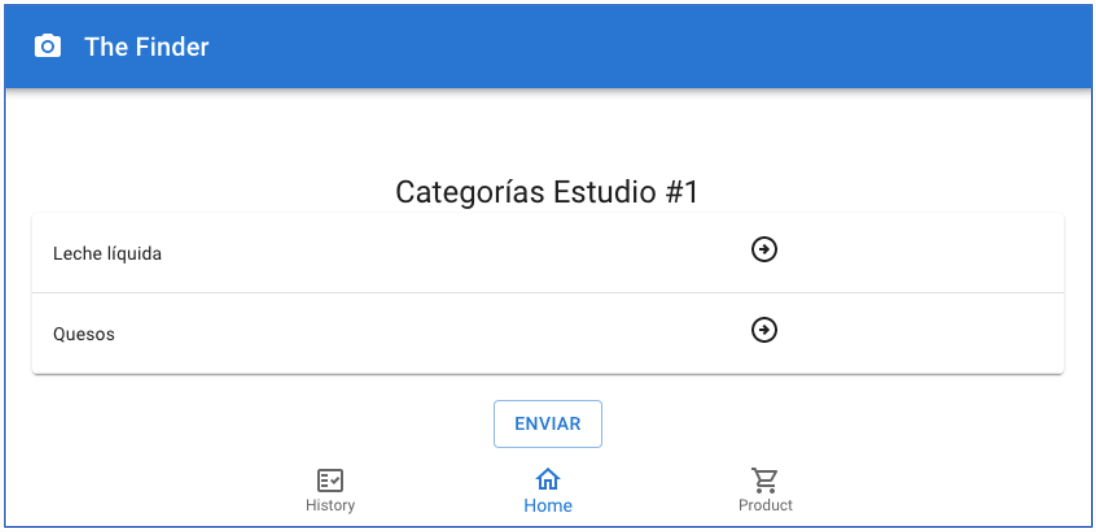

Figura 25: Interfaz Realizar estudio por categoría.

4.1.4. Interfaz capturar productos

En esta interfaz se mostrará una tabla que contendrá los productos a ser escaneados por el auditor (ver Figura 26). El auditor deberá presionar el botón de cámara para poder fotografiar la góndola de la categoría seleccionada y luego pondrá realizar estudio. La aplicación cargará la fotografía y mostrará los resultados.

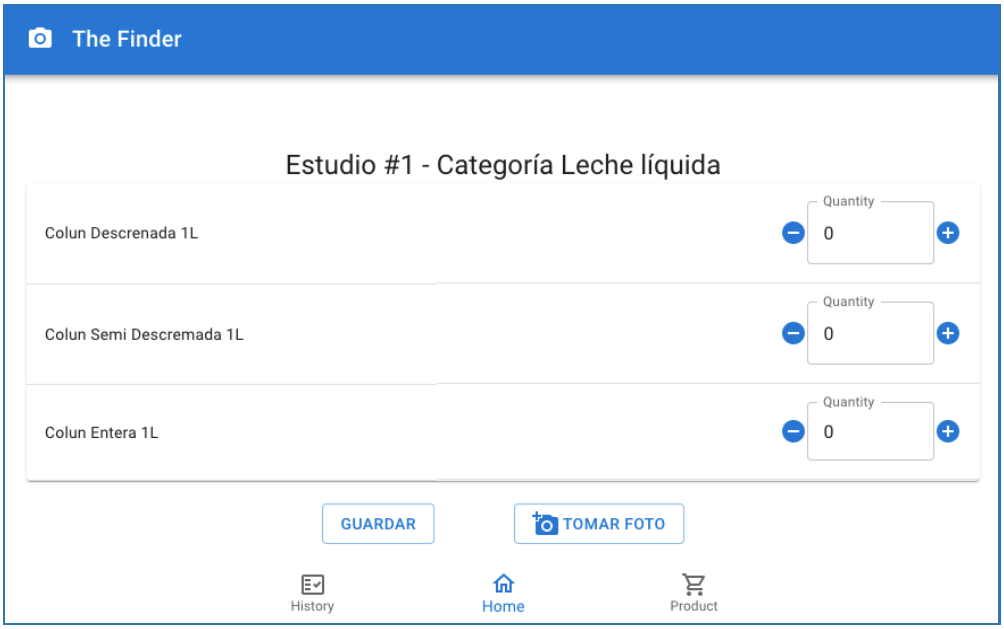

Figura 26: Interfaz Escanear productos.

4.1.5. Interfaz confirmar información de productos

El auditor podrá confirmar si la información de los productos detectados por la aplicación es correcta. Si la información no está correcta el auditor podrá corregirla inmediatamente. En la imagen de la interfaz capturar productos (figura 27) muestra la opción para editar la cantidad de productos detectados.

Figura 27: Interfaz confirmar productos

4.1.6. Interfaz para ver historial

El auditor tendrá la opción de ver el historial de los estudios realizados. Para esto tendrá que acceder a *History* y presionar el botón "detalle" como se muestra en la figura 28.

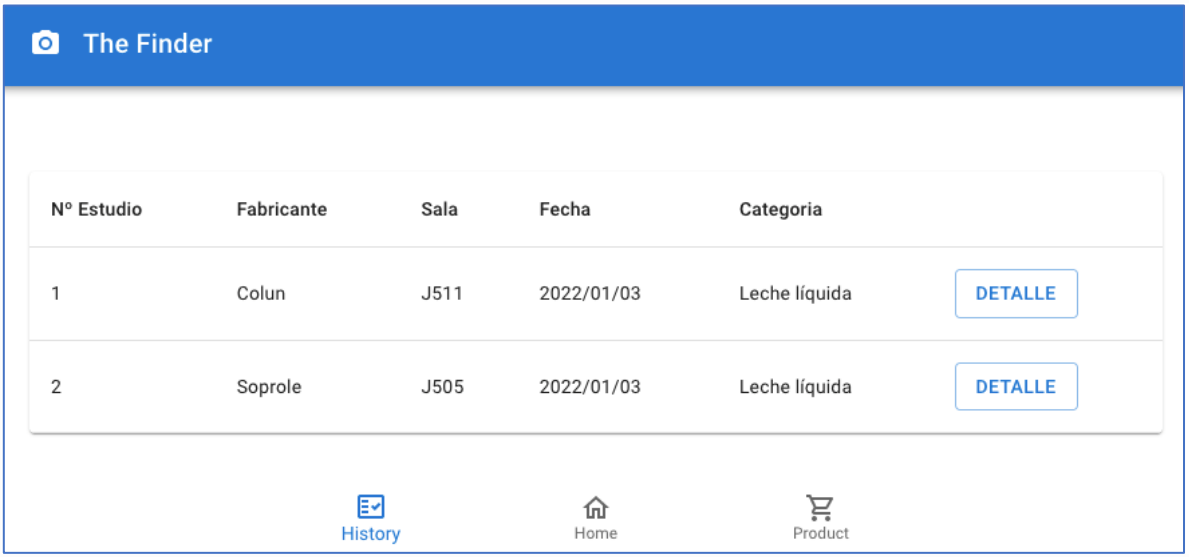

Figura 28: Interfaz ver historial

4.1.7. Interfaz cargar nuevos productos

El supervisor podrá requerir la carga de nuevos productos a la aplicación, para esto, el auditor tendrá que ingresar a *Product* y deberá seleccionar el botón "Nuevo Producto" que se muestra en la figura 29.

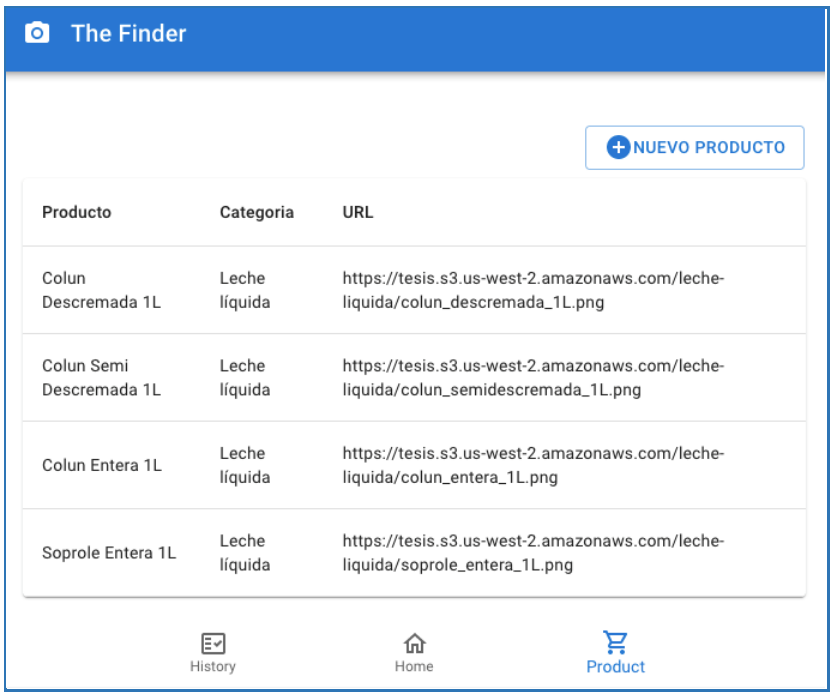

Figura 29: Administrador de productos

En esta funcionalidad el auditor podrá capturar una imagen del producto correspondiente, especificando nombre del producto, marca y tipo de producto, luego deberá presionar el botón "Registrar Producto" como se muestra en la figura 30. De esta forma, el sistema lo incorpora a sus registros para poder realizar la detección.

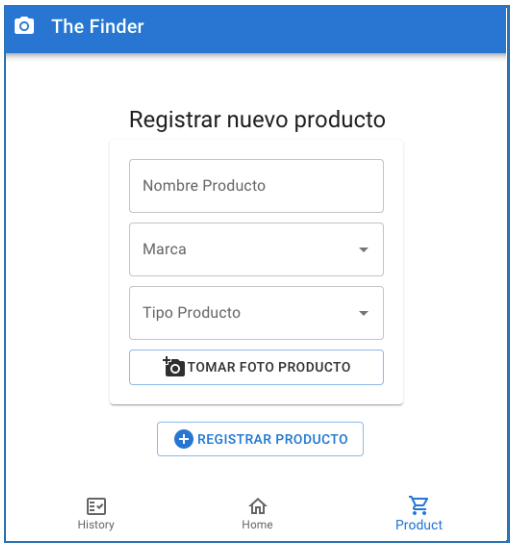

Figura 30: Interfaz cargar nuevos productos

4.2. Estimación de costos de la solución

El costo de la solución estará basado en la arquitectura definida en el capítulo 3, sección 3.3.2. *Diagrama de contenedores del sistema*. A continuación, en la Tabla 1 se presenta un detalle del estimado para la puesta en marga y primeros meses de operación.

Tipo de servicio	Componente	Región	Precio componente	Precio servicio
Amazon EC2 Service (US East (N. Virginia))				$129.81
	Compute:	US East (N. Virginia)	$121.81	
	EBS Volumes:	US East (N. Virginia)	$8	
	EBS IOPS:	US East (N. Virginia)	$0	
	EBS Throughput:	US East (N. Virginia)	$0	
Amazon S3 Service (US East (N. Virginia))				$2.89
	S3 Standard Storage:	US East (N. Virginia)	$2.3	
	S3 Standard Put Requests:	US East (N. Virginia)	$0.5	
	S3 Standard Select Data Returned:	US East (N. Virginia)	$0.07	
	S3 Standard Select Data Scanned:	US East (N. Virginia)	$0.02	
Amazon CloudFront Service				$4.78
	Data Transfer Out:	Global	$2.2	
	Data Transfer Out to Origin:	Global	$2.5	
	Requests:	Global	$0.08	
Amazon RDS Service (US East (N. Virginia))				$235.28

	DB instances:	US East (N. Virginia)	$212.28	
	Storage:	US East (N. Virginia)	$23	
AWS Support (Basic)				$0
	Support for all AWS services:		$0	
		Free Tier Discount:		$-7.71
		Total Monthly Payment:		$365.05

Tabla 1: Costo estimado de la plataforma (Fuente [10]).

El costo total de la plataforma será de 365.05 USD por mes. En esta evaluación se considera un servidor donde estará alojado el *backend*, junto con los modelos de *inteligencia artificial* y *machine learning*.

En la Tabla 2 se consideraron los costos de transferencia de datos entre los clientes y el sitio web. El costo de almacenamiento para 100GB de foto de góndolas de supermercado, además 50GB de consultas sobre los datos almacenados y 100GB de recuperación (pensando en recuperar todas las fotos). Se consideró una base de datos con 200GB de almacenamiento y 4 núcleos con 16 GB de RAM, suficiente para soportar una demanda inicial del sistema.

Capítulo 5: Evaluación de la solución

En este capítulo se muestra en detalle cómo se llevó a cabo el proceso de evaluación de la solución y cuáles fueron los resultados obtenidos a través de los experimentos realizados en la categoría de leches líquidas. Finalmente, se explicarán las limitaciones de la evaluación y los problemas encontrados.

5.1. Proceso de evaluación

Para comprender los resultados de los experimentos es necesario determinar qué puntos de control de calidad o evaluaciones serán determinantes.

5.1.1. Métodos de evaluación

A continuación, se listan los métodos de evaluación y una descripción que permitirá comprender los criterios aplicado en los experimentos.

Evaluación de los cluster de imágenes. Utilizaremos el coeficiente de silueta para comprender la superposición de los clusters y elbow (codo) para validar si el número de clusters está relacionado con la cantidad de agrupaciones realizadas.

Coeficiente de silueta (silhoutte score). El coeficiente de silueta se calcula utilizando la distancia media dentro de un grupo A y la distancia media del grupo B más cercano. Permite aclarar que la distancia entre una muestra del grupo A y el grupo B más cercano del que la muestra no forma parte. Donde, el mejor valor será 1 demostrando que la muestra del Grupo A no corresponde al grupo B (se encuentran distante), y el peor valor será -1 que podría indicar que una muestra fue asignada al grupo equivocado. Para los valores cercanos a 0 significa que los clusters están superpuestos.

Método del codo (Elbow curve). El método del codo será de ayuda para determinar el número de centroides que debe tener nuestro cluster, dado que al principio estamos agrupando por un número dado de marcas o tipos de productos pertenecientes a una marca, es necesario validar que el número de centroides esté relacionado con la cantidad de tipos o marcas. El método del codo utiliza la suma de las distancias al cuadrado de las muestras a su centro del grupo más cercano. Entonces, definimos un total de centroides a comprobar e iteramos desde cero a N que será el número por comprobar.

Evaluación del sistema. Se considerará el tiempo completo en realizar el reconocimiento de los productos, además del nivel de acierto con los resultados.

5.2. Resultados obtenidos en la detección de productos

Para evaluar la solución se realizaron tres experimentos: reconocimiento de marcas, reconocimiento de tipo de producto por marca, y un experimento final que consistió en realizar el proceso completo que realiza un auditor.

5.2.1. Experimento categoría leches líquidas, reconocimiento de marcas

El siguiente experimento tiene por objetivo comprobar la factibilidad de clasificar marcas desde un aprendizaje auto-supervisado.

Limpieza de imágenes
Para llevar a cabo este experimento se debe realizar una limpieza del conjunto de datos (explicado en conjunto de datos del capítulo 3, sección 3.6), de las cuales quedaron **579 imágenes** de 601 imágenes correspondiente a la categoría de leches líquidas. Los criterios de limpieza fueron los siguientes:

- La imagen no debe presentar movimiento durante su captura.
- Debe poder verse la marca del producto.
- La luminosidad de la foto debe al menos permitir reconocer la marca y evitar que la luz de la sala no oculte con el brillo la marca.
- El ángulo de captura de la imagen debe ser frontal.

Detectar productos en la góndola
Para hacer uso del modelo *Precise Detection in Densely Packed Scenes* se utilizaron las configuraciones por defecto especificadas por el autor. En cuanto a los parámetros personalizados para el experimento, solo se cambió el nivel de precisión a un 70%. El proceso para detectar los productos fue el siguiente:

- Se realizó una copia de las 579 imágenes de la categoría de leche líquida en la carpeta que lee el modelo.
- Se configura la precisión al 70%
- Se inicia la detección de los productos obteniendo 3.685 filas de anotaciones con las coordenadas (x1, x2, y1, y2) de píxeles de los productos.

Como resultado, en las 579 imágenes, se presume haber detectado 3.685 productos.

Entrenar extractor de características
Para obtener las características representativas de cada imagen, se debe utilizar el modelo *bootstrap your own latent* (byol), pero antes debe ser entrenado con la categoría de leche líquida. Primero debemos cortar los productos detectados, y separarlos en carpetas como *train*, *val*, *test*. En la Tabla 2 se muestran los resultados obtenidos.

Nombre carpeta	Cantidad de imágenes
train	2.579
val	737
test	369

Tabla 2: Cantidad de imágenes por carpeta

Para entrenar el modelo se utilizó la carpeta de entrenamiento (*train*) con 2.579 productos utilizando los siguientes híper-parámetros:

- Épocas: 200
- Tasa de aprendizaje (Learning Rate): 0.0003 (3e-4)
- Tamaño del lote (batch size): 32
- Tamaño de imagen: 256

El tiempo de entrenamiento del modelo utilizando una *graphics processing unit (GPU)* marca *Nvidia* modelo *Tesla P100* de 16GB de capacidad de memoria fue de 16 horas.

Realizar cluster de marcas

Para el *clúster* se utilizó un subconjunto de marcas tales como Colun, Líder y Soprole. utilizando los parámetros por defecto especificados para el modelo de *k-means*. En la figura 31 se muestra los *cluster* generados para las marcas Colun, Líder y Lonco leche. Se puede apreciar que Lonco leche queda muy alejada en comparación entre Colun y Líder.

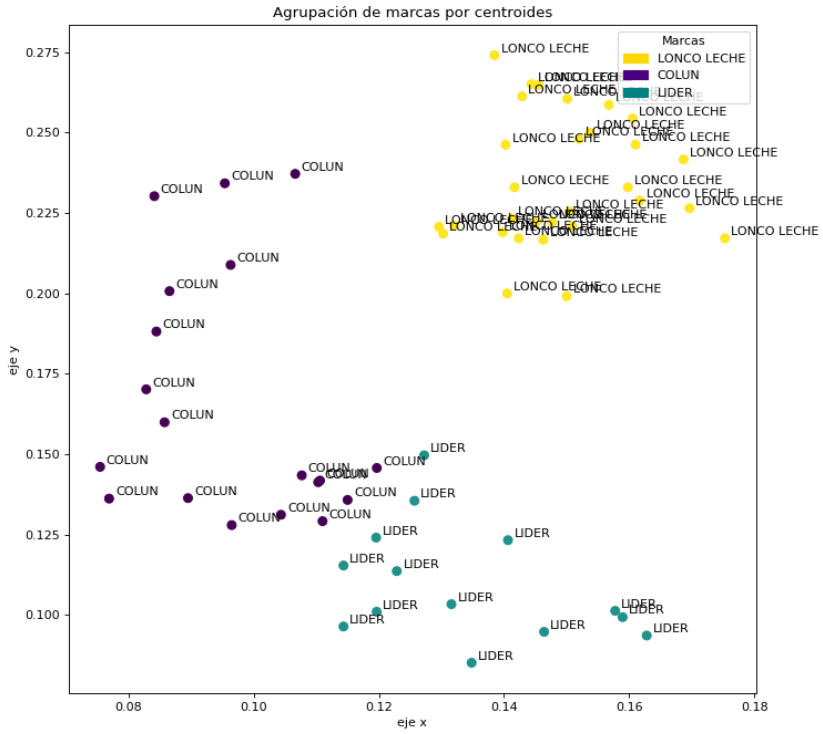

Figura 31: Gráfico de grupo de marcas

Como se muestra en la Tabla 3, los resultados entre Líder y Colun se encuentran mezclados, no logró diferenciar todas las muestras con precisión.

Marca	Cantidad etiquetadas	Cantidad predicción	Porcentaje de error
COLUN	16	18	12.5%
LÍDER	17	15	11.8%
LONCO LECHE	30	30	0%

Tabla 3 Resultados de las agrupaciones realizadas por el cluster k-means.

Como resultado en la Tabla 3 observamos que Líder tuvo una falla en la detección, se esperaba un total de 17 etiquetas correctas y se obtuvieron 15, lo cual significa que la frontera entre los dos grupos de marca (Colun y Líder) no está bien delimitada. Para complementar este análisis se obtuvo el *coeficiente de silueta (silhouette score)* con un valor de 0.4668, el cual permite entender si los grupos se encuentran superpuestos, entre más cerca del valor 1 significa que no encuentran superpuestos (esto se mostrará en la Figura 32).

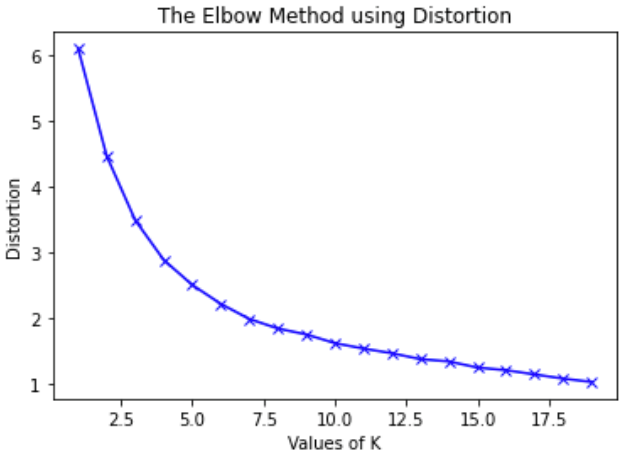

Figura 32: Método del codo para validar que el número de centroides es adecuado para el grupo de marcas

En la Figura 32 se puede observar que no hay un quiebre notorio en el número de *centroides*. Esto puede significar que el número de grupo no tiene un horizonte determinado para denominarlo "grupo de marcas". A partir de los resultados obtenidos podemos destacar lo siguiente:

- Fueron detectados 3.685 productos con una precisión del 70%.
- El tiempo de entrenamiento del modelo BYOL fue de 16 horas.
- La reducción de imágenes por limpieza fue del 3.66%.

5.2.2. Reconocimiento de tipo de productos por marca

Para el siguiente experimento hemos utilizado el mismo conjunto de imágenes del *experimento categoría leches líquidas, reconocimiento de marcas*, y también el mismo extractor de características previamente entrenado. La diferencia está en poner a prueba la agrupación de los tipos de productos por una marca que será Colun. Para ello, se llevaron a cabo los siguientes pasos:

Realizar cluster de tipo de productos

Se seleccionó una marca en específico, en este caso Colun para ver de qué forma se realiza la separación de los tipos de productos. Se utilizarán tipos como Entera, Descremada y Semi Descremada para comprobar que el extractor de características tiene una aplicación en esta segunda parte del proceso. Además, comprender cómo se distribuyen los tipos de productos en el gráfico de dispersión de la Figura 33.

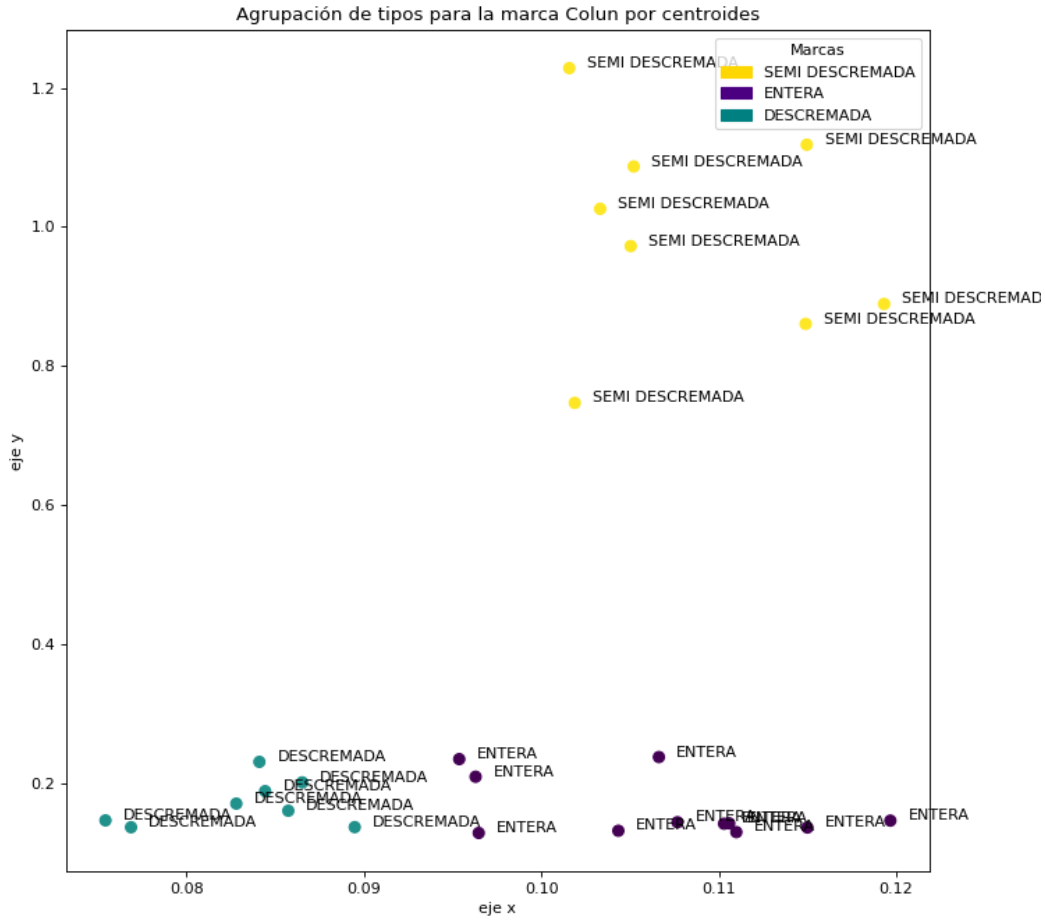

Figura 33: Grupo de tipo de productos relacionados con la marca Colun.

En la figura 33 se muestra los *cluster* generados para los tipos de leches Entera, Descremada y Semi Descremada de la marca Colun. Se puede observar que el tipo Semi Descremada queda lejos de los tipos Descremada y Entera, nuevamente podemos suponer que el horizonte de los tipos puede ser poco decisivo. Utilizando el coeficiente de silueta obtenemos 0.4430 que ayuda a comprender que tan consolidado se encuentra un centroide respecto a las muestras.

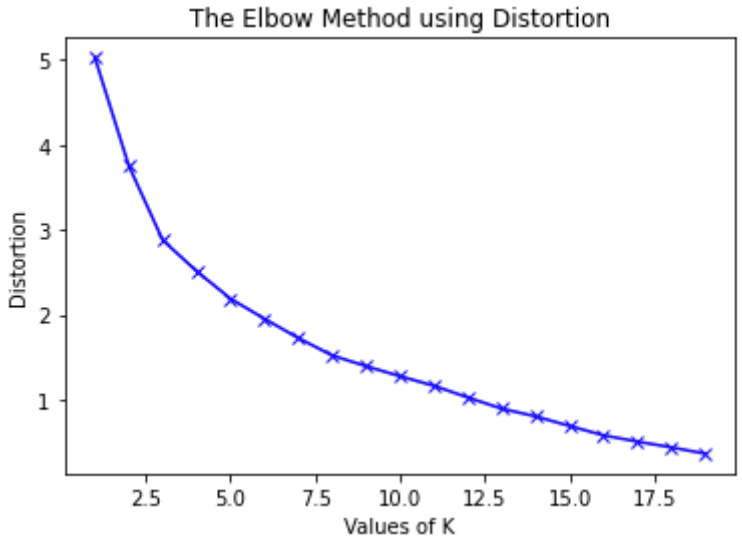

Figura 34: Método del codo para validar que el número de centroides es adecuado para el grupo de tipo de productos de la marca Colun.

En la Figura 34 se muestra que existe un ligero quiebre en el número 3, esto está alineado con el número de tipo de productos correspondiente a la marca Colun. También, podemos notar que falta determinación en el quiebre del método del codo. A partir de los resultados obtenidos podemos destacar lo siguiente:

- Comprobamos que el extractor de características si realiza una separación apropiada de los tipos de productos.
- Validamos que 3 *centroides* se encuentran relacionados con sus respectivos tipos de productos.

5.2.3. Experimento de flujo completo

En el experimento de flujo completo realizamos una prueba completa de la cadena de procesos que involucran los modelos. Para ello, se realizará la prueba con una imagen de góndola de supermercado de la cual conozcamos sus etiquetas y evaluaremos los resultados que entrega el sistema. Para comprender a grandes rasgos cuál será la cadena de procesos, a continuación, se detallan los pasos que se siguieron:

1. El *auditor* toma una foto de la góndola de supermercado.
2. Se envía la imagen al sistema de reconocimiento.
3. Se detectan los productos en la góndola.
4. Se cortan individualmente los productos.
5. Se extraen las características relevantes de los productos.
6. Se utiliza el primer *cluster* de marcas para determinar la marca de los productos.
7. Según la marca de los productos, esto se envían a sus *clusters* correspondientes. Se utiliza el *cluster* de tipo de productos por marca.
8. Se entregan los resultados ya procesados al *auditor*.

En el experimento obviaremos los pasos N° 1, 2 y 8, que corresponden a la recepción y entrega de información del auditor, dado que queremos comprender la precisión del sistema.

Figura 35: Foto de góndola de supermercado sin procesar a la izquierda. Foto de góndola de supermercado procesada por el detector de productos a la derecha.

Utilizaremos la imagen que se observa en el lado izquierdo de la Figura 35, y después veremos cómo se detectan los productos en la figura que se muestra en el lado derecho.

Luego de analizar la imagen el modelo detecta productos y falsos positivos, en la Figura 35 se marcan los packs de 12 leches ubicados en la parte inferior de la góndola como un producto individual, de esta forma se detectaron 107 posibles productos. Dada su similitud con envases individuales que no deberían ser considerados para los resultados que busca el *auditor*, estos se excluyen del experimento al aplicar un filtro de 80% de precisión como parámetro del modelo, quedando un total de 53 imágenes individuales que corresponde al proceso Nº 3 y Nº 4. En la Tabla 4 se muestra las marcas que fueron encontradas con su respectiva cantidad y porcentaje de detección en la góndola.

Marca	Cantidad de productos góndola	Cantidad de productos detectados	Porcentaje de detección en góndola
Colun	37	32	86%
Soprole	19	13	68%
Surlat	8	8	100%

Tabla 4: Detalle de cantidad de productos por marca detectados.

La extracción de características demoró un total de 10 segundos para 53 productos, que corresponde al proceso Nº5. Como resultado obtenemos una lista de vectores de características que pueden ser agrupados en el *cluster* de marcas que será utilizado por el proceso Nº 6.

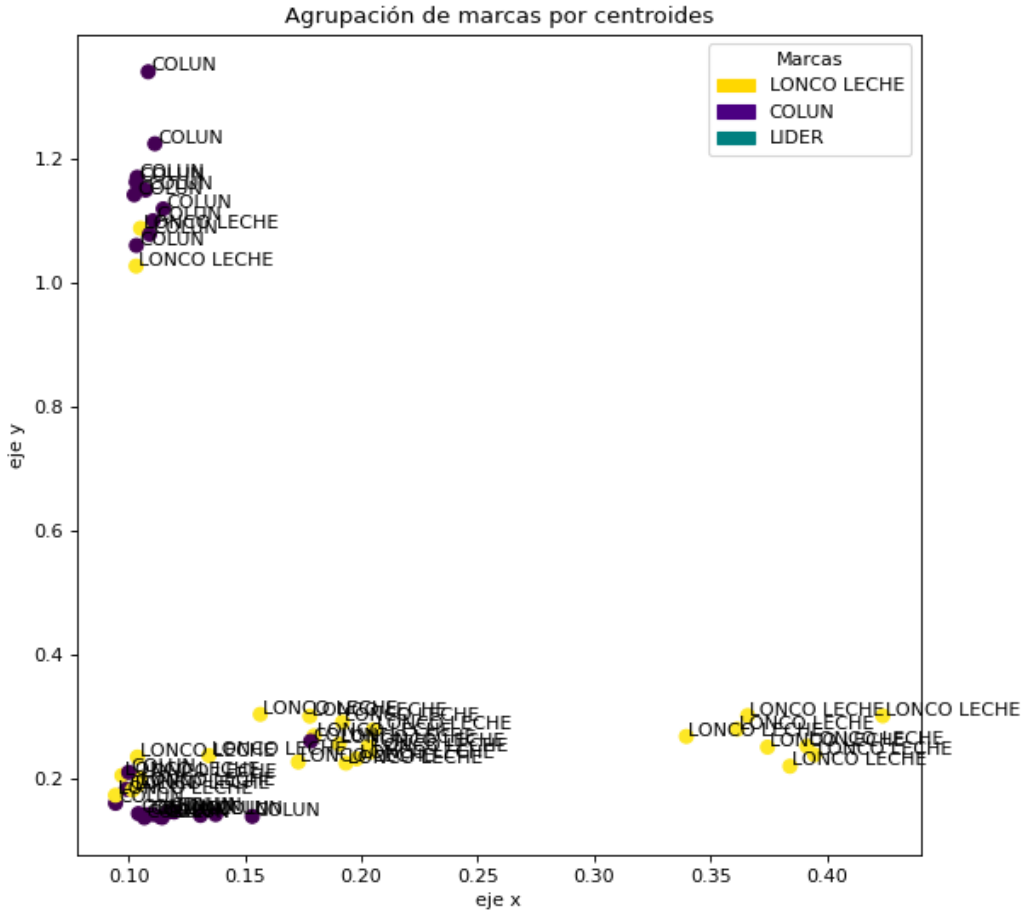

Figura 36: Agrupación de marcas por centroides, experimento flujo completo.

Como se muestra en la Figura 36, hay marcas que no conoce el *cluster* dado que no fue entrenado para identificarlas, lo cual significará que estarán mal etiquetadas. En la tabla 5, se muestra cómo Soprole y Surlat fallan en su identificación, dado que no conoce la marca.

Marca	Total productos	Correcta predicción	Incorrecta predicción	Porcentaje de detección
Colun	32	24	8	75%
Soprole	13	0	13	0%
Surlat	8	0	8	0%

Tabla 5: Predicciones para la marca colun.

En la tabla 6 se puede ver que Colun siendo una marca conocida por el sistema tuvo un 75% de aciertos con 24 etiquetas correctas y 8 incorrectas de un total de 32 productos. A continuación, se utilizará la marca Colun para realizar la predicción de tipos de productos.

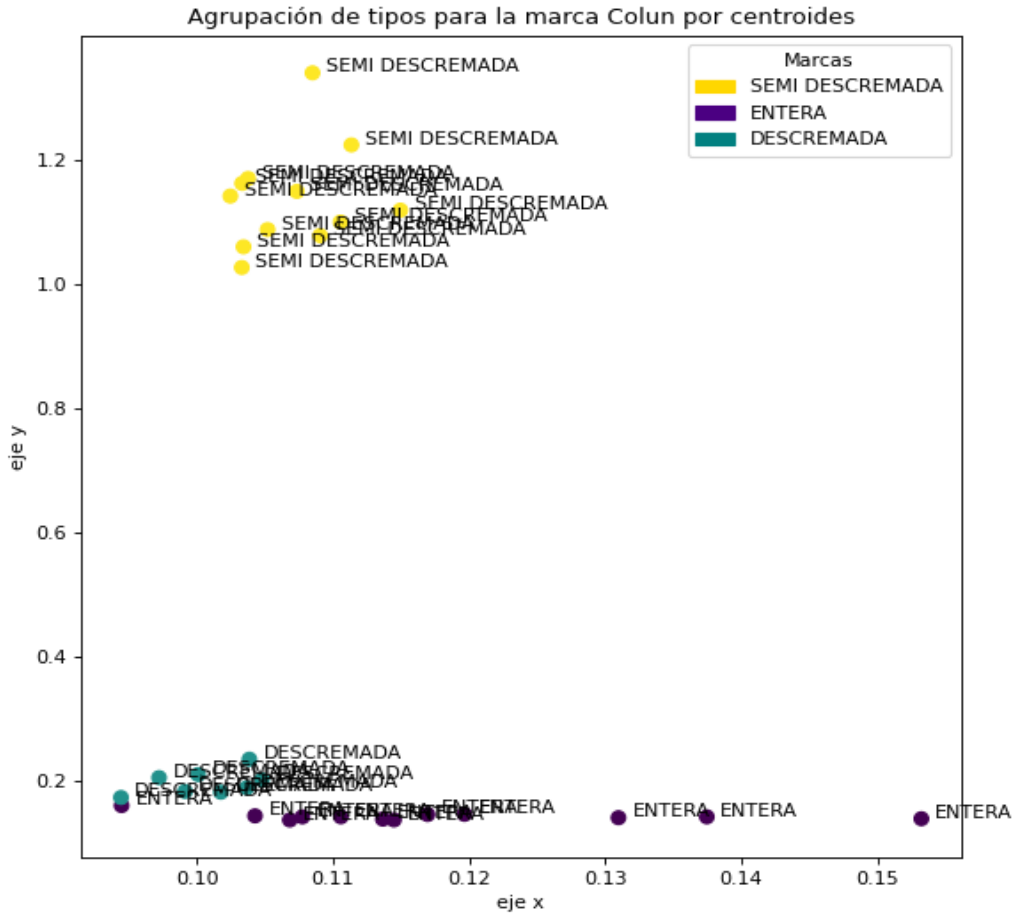

Figura 37: Grupo de tipo de productos para la Marca Colun en el proceso de flujo completo.

En la Figura 37 se muestra cómo los tipos de productos Descremada y Entera tiene una posible superposición entre sus grupos, sin embargo, en la tabla 6 que muestra la evaluación de la cantidad de aciertos sobre las predicciones vemos que el 100% de los tipos de productos tienen su correcta etiqueta.

Tipo de producto	Total productos	Correcta predicción	Incorrecta predicción	Porcentaje de detección
Entera	12	12	0	100%
Semi Descremada	12	12	0	100%
Descremada	8	8	0	100%

Tabla 6: Predicciones para los tipos de productos de la marca Colun

La correcta identificación de los tipos puede significar que los grupos individuales por marca se especializan mejor que un modelo genérico por marca. También podemos notar que la identificación de tipo de productos tiene un 100% de acierto en comparación con el modelo que identifica marcas que obtiene un 75% de acierto.

Durante el experimento se calcularon los tiempos de los procesos principales, los cuales se muestran en la tabla 8.

Proceso	Tiempo en segundos
Tiempo en detectar productos (proceso Nº 3)	20
Tiempo en cargar modelo BYOL (proceso interno)	2
Tiempo en cortar imágenes individualmente (proceso Nº 4)	9
Tiempo en extraer características para 53 productos (proceso Nº 5)	10
Tiempo en clasificar por marca (proceso Nº 6)	0.9
Tiempo en clasificar por tipo de producto (proceso Nº 7)	0.1
Tiempo total	**42**

Tabla 7: Tiempo que requiere el sistema en realizar los procesos para el reconocimiento de los productos.

En la tabla 7 se muestra el tiempo total que lleva al sistema en procesar una imagen de una góndola de supermercado, siendo este un total de 42 segundos por imagen. A partir de los resultados obtenidos podemos resaltar lo siguiente:

- El cluster de marcas tiene un 75% de acierto sobre la marca Colun.
- El cluster de tipos de productos tiene un 100% de acierto sobre la marca Colun.
- Vemos que una posible superposición de cluster puede suceder y obtener una correcta predicción.
- El tiempo que demora el sistema el reconocer un grupo de marcas con sus respectivos tipos desde una foto de una góndola de supermercado es de 42 segundos.

5.3. Limitaciones de la evaluación

En la realización de los experimentos encontramos limitaciones que deben ser consideradas para el uso de esta solución.

Problema de imágenes desenfocadas o de baja calidad. Durante el desarrollo de los experimentos se tuvo que realizar una revisión minuciosa (producto por producto) de la nitidez de cada muestra. Si una muestra se encuentra desenfocada o tiene un rastro de movimiento durante la captura de la imagen, puede afectar los resultados. La principal limitación encontrada es la falta de rigurosidad en el proceso de captura de imagen puede afectar los resultados.

Problema con la similitud de productos. Como se puede ver en los resultados de las agrupaciones por marca y tipo de producto, existe un problema con la similitud de productos, es por esto que será necesario mantener los grupos pequeños para realizar un correcto seguimiento de los clusters generados. La limitación se encuentra en generar clusters con demasiados candidatos como marcas o tipos de productos.

Problema en la variabilidad de dispositivos. Durante la limpieza de los datos se identificaron distintos dispositivos móviles utilizados para capturar fotos de góndolas de supermercado, es posible que la variabilidad de dispositivos móviles altere la extracción de características y afecte el reconocimiento de los productos. La limitación se relaciona con una variabilidad considerable de dispositivos consultando por reconocimiento de productos.

Capítulo 6: Conclusiones y trabajo a futuro

En la presente tesis se abordó el problema de automatizar un proceso de levantamiento de información que realizan los *auditores* al momento de completar un *estudio*. Como solución se realizó un sistema de reconocimiento de productos utilizando aprendizaje auto-supervisado permitiendo levantar cantidades de productos por marcas y tipos de productos de forma escalable.

Un estudio de presencia de 200 productos puede tardar 70 minutos aproximadamente en levantar la información. Por otro lado, sabemos que el sistema tarda 42 segundos en procesar los productos de una imagen. Ahora si realizamos una comparación entre el estudio de presencia y el sistema, obtenemos lo siguiente:

- Si tenemos un caso donde hay grupos de 3 productos por cada imagen, el tiempo que tarda el sistema es de 46 minutos, en comparación con los 70 minutos de un estudio de presencia. Con esto, encontramos una reducción de un 34% en el tiempo de levantamiento de información.
- Si tenemos un caso donde hay grupos de 2 productos por cada imagen, el tiempo que tarda el sistema es de 70 minutos, en comparación con los 70 minutos de un estudio de presencia. Con esto, se iguala el tiempo del auditor.
- Por otro lado, si consideramos un peor caso, donde por cada imagen encontramos solo un producto de la lista, obtenemos un tiempo total del estudio de alrededor de 140 minutos en levantar la información de 200 productos en el sistema, comparado con los 70 minutos que tarda un estudio de presencia. Con esto, obtenemos un aumento del tiempo en un 100%, lo que significa que para un caso como el que se mencionó antes, el sistema demora el doble del tiempo.

Respecto a los casos expuestos, podemos determinar que dado una lista de productos que estén en una misma categoría (posiblemente agrupados) el sistema es una opción viable para ser utilizado, sin embargo, dado el caso en que los productos no estén relacionados en una misma categoría, podemos dejar fuera el sistema de reconocimiento de productos y considerar como mejor opción que el auditor realice el estudio manualmente.

Dentro de los desafíos planteados durante la tesis, se resuelven problemas como escalar un sistema de reconocimiento sin incurrir en gastos exponenciales debido al aumento de productos, tipos de productos o categorías. Como también, se tuvo que resolver capturas de góndolas densamente pobladas por productos, posiciones variadas de envases, intra-variabilidad de una marca y el desafío de reconocer sin requerir de un número elevado de muestras etiquetadas.

En los experimentos realizados para determinar el comportamiento de los modelos observamos que con un 80% de precisión podemos detectar productos en poblaciones densas de una imagen de góndola de supermercado, además la precisión de los *clusters* generados por el algoritmo de *k-means* es de 75% para las marcas y un 100% para los tipos de productos, esto significa que es podemos distinguir un grupo pequeño de producto siguiendo una estrategia de dividir y conquistar. Estos precedentes revelan que es posible llevar a cabo un proceso auto-supervisado de clasificación de productos.

En cuanto al costo del sistema la estimación es variable según sea la cantidad de marcas y tipos que se requiera dar de alta, sin embargo, un costo estimado por mes es de USD 365.05 por mes. Lo

cual justifica la factibilidad de implementar el sistema de reconocimiento como apoyo en el relevamiento de información que realiza un *auditor* durante los *estudios*.

Como beneficio que se pueden obtener del sistema de reconocimiento de productos es el respaldo de los *estudios* realizados por los *auditores* y poder entregarlos a los *fabricantes* a modo de informe. Además, volver a consultar la información con una referencia en historial de *estudios*.

En lo que respecta a trabajos futuros, existe la posibilidad de escalar el sistema y mejorar la velocidad de procesamiento de las imágenes de góndolas de supermercado, dado que en los experimentos se utilizó una CPU convencional en lugar de una GPU que si mejora el rendimiento de los modelos.

Bibliografía

[1] Vinyals, Oriol, Blundell, Charles, Lillicrap, Timothy, Kavukcuoglu, Koray, Wierstra, Daan (2016). *Matching Networks for One Shot Learning. Advances in neural information processing systems, Volume 29*. NIPS, 2016.

[2] Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey E. (2012). *ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS, 2012.

[3] Alom, Md. Zahangir & Taha, Tarek & Yakopcic, Christopher & Westberg, Stefan & Hasan, Mahmudul & Esesn, Brian & Awwal, Abdul & Asari, Vijayan. (2018). *The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches*. ArXiv, 2018.

[4] ImageNet website. *Inicio, registros de imágenes y clases disponibles en el conjunto de datos*. (2021). ImageNet. https://image-net.org. Ultimo acceso: 29-03-2022.

[5] Zhang, Zhao & Hsieh, Cho-Jui & Demmel, James & Keutzer, Kurt. (2018). *ImageNet Training inMinutes.* ICPP, 2018.

[6] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. (2009). *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations.* ICML 2009.

[7] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). *ImageNet Large Scale Visual Recognition Challenge*. International Journal of Computer Vision, 2015.

[8] Zheltonozhskii, E. (2020). Self-Supervised Learning for Large-Scale Unsupervised Image Clustering. ArXiv, 2020.

[9] Amazon Rekognition. *Análisis de imágenes y videos con machine learning*. (2022). Amazon Web Services. https://aws.amazon.com/es/rekognition. Ultimo acceso: 13-01-2022.

[10] Amazon Web Services. *Amazon Web Services Simple Monthly Calculator*. (2022). Amazon Web Services. https://calculator.aws. Ultimo acceso: 19-04-2022.

[11] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, Michal Valko. (2020). *Bootstrap your own latent: A new approach to self-supervised Learning, Volume 3.* NIPS, 2020.

[12] Eran Goldman, Roei Herzig, Aviv Eisenschtat, Oria Ratzon, Itsik Levi, Jacob Goldberger, Tal Hassner. (2019). *Precise Detection in Densely Packed Scenes, Volume 3*. IEEE, 2019.

[13] C4 Model. *The C4 model for visualising software architecture. (2022).* C4 Model. https://c4model.com. Ultimo acceso: 20-01-2022.