

# Proyecto

Reconocimiento de objetos en góndolas de supermercados

Integrantes: Bastián Garcés G.  
María José Liberona T.  
Profesor: Claudio A. Pérez  
Auxiliar: Jorge Zambrano I.  
Ayudante de laboratorio: Juan Pérez C.

Fecha de entrega: 7 de abril 2022  
Santiago de Chile

# Índice de Contenidos

<b>1. Ante-proyecto</b>	<b>1</b>
1.1. Problemática . . . . .	1
1.2. Metodología . . . . .	1
1.2.1. Deep Learning for Retail Product Recognition: Challenges and Techniques .	1
1.2.2. Grocery product detection and recognition . . . . .	1
<b>2. Resultados esperados</b>	<b>2</b>
<b>Referencias</b>	<b>2</b>

# 1. Ante-proyecto

## 1.1. Problemática

La visión computacional o también conocido como visión por computador es un concepto que ha adquirido gran relevancia en los últimos años, debido a la gran cantidad de aplicaciones que posee como son: detección de defectos, reconocimiento de rostros, seguimiento de movimientos, entre otras. Una de las principales tareas de la visión computacional está enfocado en el reconocimiento, ya sea de personas, animales u objetos. Teniendo en consideración lo anterior, se ha decidido realizar el proyecto del curso EL7007 enfocado en el reconocimiento de objetos en góndolas de supermercados, donde se espera implementar un algoritmo capaz de discernir entre distintos tipos de productos.

## 1.2. Metodología

Para la realización de este proyecto se proponen dos metodologías estudiadas en investigaciones anteriores, específicamente, en los artículos *A deep learning pipeline for product recognition on store shelves* [1] y *Grocery product detection and recognition* [2]. Para llevar a cabo el proyecto se hará uso de la Base de Datos *GroZi-120* [3], que provee de imágenes y vídeos para la detección de 120 productos diferentes.

### 1.2.1. Deep Learning for Retail Product Recognition: Challenges and Techniques

En este caso, el procedimiento de detección y reconocimiento de imágenes consta de 3 partes:

1. La primera parte consiste en la implementación de un sistema de detección mediante una CNN, la cual se encarga de extraer propuestas de región de la imagen de consulta. Para ello, se propone como red de detección la red *YoloV2*<sup>1</sup>, ya que garantiza el rendimiento en tiempo real en una GPU y la disponibilidad de la implementación original.
2. Luego, cada propuesta detectada se recorta de la imagen de consulta y se ingresa a otra CNN (Embedder) que calcula una representación de la imagen. Para la red principal de Embedder se propone la red *VGG-16*<sup>2</sup>, la cual se encuentra pre-entrenada en la tarea de clasificación *Imagenet-1000*. Cabe mencionar que el proceso de entrenamiento de Embedder se lleva a cabo utilizando las imágenes de referencia de los productos.
3. Por último, se plantea realizar el reconocimiento del producto a través de una búsqueda de similitud KNN en una base de datos de representaciones pre-calculadas fuera de línea por el Embedder en las imágenes de referencia.

### 1.2.2. Grocery product detection and recognition

Para el caso del presente artículo la metodología se divide en preselección, selección fina y post-procesamiento. A continuación se presenta la labor de cada una de las fases:

<sup>1</sup> <https://github.com/pjreddie/darknet>

<sup>2</sup> <https://github.com/tensorflow/models/tree/master/research/slim>

1. La etapa de preselección está destinada a seleccionar el conjunto inicial de ventanas candidatas sobre la información conjunta obtenida de la posición de las esquinas y la distribución de color de las imágenes en cuestión. Para la detección de esquinas se propone el uso de un detector de esquinas Harris, luego para cada esquina detectada se proponen 4 ventanas candidatas. Posteriormente se evalúa en cada una de estas ventanas candidatas la varianza de la escala de grises en la ventana en cuestión y el porcentaje de píxeles en primer plano. Luego, si se cumplen ambas condiciones, se calcula el histograma de color 3D calculado en el espacio de color YCbCr, se compara con el histograma de referencia del producto destino mediante la intersección de histogramas habiendo realizado previamente una normalización del color en las imágenes de entrenamiento y prueba, lo cual se puede lograr mediante el uso de un algoritmo CLAHE. Posterior a la intersección, se elige la ventana candidata cuyo histograma tiene una mayor intersección con el producto objetivo y dicha ventana pasa al postprocesamiento.
2. En la etapa de selección fina se aplican características más robustas para la selección de candidatos. Es en este punto en que se proponen dos métodos para la obtención de vectores de características, los que corresponden al algoritmo de Bolsa de palabras (Bag of Words) y una Red Neuronal Convolutiva Profunda. El algoritmo de bolsa de palabras se utilizaría para obtener vectores de características para las imágenes de entrenamiento, dichos vectores son independientes de la distancia entre los objetos y la cámara. Luego para el caso de la red convolutiva se utilizaría la red *AlexNet* ingresando imágenes reescaladas a 227x227 píxeles, en el artículo también se usan imágenes en escalas de grises para poder comparar el algoritmo de Bag of Words con la CNN (para ponerlos en igualdad de condiciones), luego se comparan los puntajes del vector obtenido de la red con los de referencia y se conservan aquellos con mayor score. Sin embargo, es necesario un postprocesamiento para elegir un único resultado como factible.
3. En la etapa de postprocesamiento se utilizan técnicas para filtrar los resultados, entre estas técnicas está el NMS o clustering.

## 2. Resultados esperados

Para la evaluación del desempeño del proyecto, se pretende utilizar las métricas *Recall* y *Precision*, donde se espera obtener un gran porcentaje de reconocimientos correctos en las fases de prueba.

## Referencias

- [1] A. Tonioni, E. Serra, and L. Di Stefano, “ArXiv:1810.01733v3 [CS.CV] 27 Jan 2019,” 27-Mar-2019. [Online]. Available: <https://arxiv.org/pdf/1810.01733.pdf>. [Accessed: 28-Mar-2022].
- [2] A. Franco, D. Maltoni, and S. Papi, “Grocery product detection and recognition,” Expert Systems with Applications, 21-Mar-2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417417301227>. [Accessed: 28-Mar-2022].
- [3] “120 database,” GroZi. [Online]. Available: <http://grozi.calit2.net/grozi.html>. [Accessed: 31-Mar-2022].