

# Products-6K: A Large-Scale Groceries Product Recognition Dataset

## 1 DATASET CHARACTERISTICS

The Products-6K dataset is characterized by its large scale in terms of unique SKUs, as it provides 12.917 candidate images associated with 6.348 unique SKUs and 373 query images linked with 104 unique SKUs. The available images of both the candidate and the query set include single-product illustrations, with the former being captured in laboratory conditions and the latter being acquired in a supermarket environment. This partition was driven by the main goal of our dataset, which is to rely on an existing product’s database that most supermarkets generate in laboratory conditions, so as to facilitate the product recognition task in a realistic supermarket scenario. Moreover, in an effort to mimic such a realistic scenario, the query images were acquired from several viewpoints under various conditions in terms of lighting and clutter, considering that the product has been removed from the shelf and its being held by the user in his/her hand. The different viewpoints and the various lighting/clutter levels are in essence the two major challenges for the efficient product recognition proposed dataset. Finally, each image (of both partitions) is accompanied by textual annotation in various levels of detail, including the product’s SKU, its specific and broad product category and its brand.

The dataset can be accessed via Zenodo<sup>1</sup>, where visual and textual descriptors for the entirety of the candidate and query sets are available. For reasons of corporate interest, the candidate and query images of 500 and 104 unique SKUs are available, but the interested reader is encouraged to follow the contact details for requesting access to the full dataset.

### 1.1 Construction of the candidate image set

The set of candidate images was provided by the Greek supermarket Masoutis<sup>2</sup>, with a total of 12.917 product images being available, corresponding to 6.348 unique product SKUs. Each image associated with a specific SKU provided a different product view, with an example illustrated in Figure 1a. The frontal view was available for all SKUs, while the availability of other views (e.g. left/right side view) varied among SKUs.

Each product image was directly linked with the product’s SKU and was accompanied by textual information corresponding to different levels of detail, starting from broad descriptions and ending up to a specific product description. The first annotation level provides information about the product’s broad category (BPC), the second describes the product’s specific category (SPC), the third determines its brand category (BC), while the last level presents the exact product description (PD). For example, for Product SKU 18283 (Fig. 1a), the product’s BPC would be *Coffee*, its SPC would be *Filtered Coffee*, its BC would be *Jacobs* and its PD would be *Coffee Jacobs Vanilla 250 grams*.

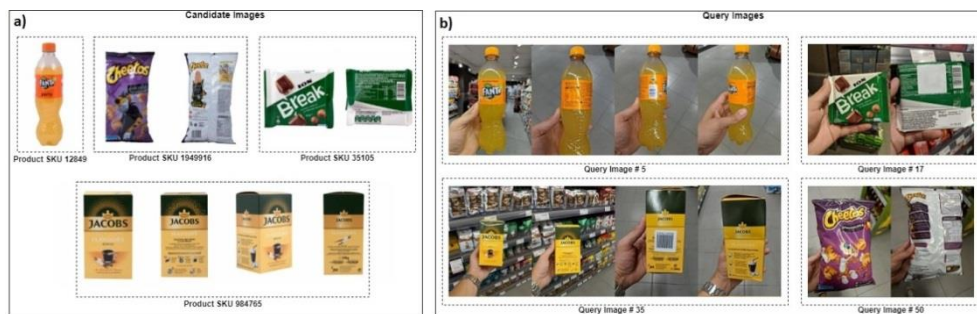


Figure 1 Sample images from (a) the set of candidate images and (b) the set of query images.

To further enhance the available textual information we employed the optical character recognition (OCR) mechanism provided by the Google Cloud Vision API<sup>3</sup> as a means to extract the accessible text (in both Greek and English) from the product packages. The acquired textual information for each product was concatenated with the product’s PD, resulting in an enhanced

<sup>1</sup> Products-6K: A Large-Scale Groceries Product Recognition Dataset - <https://zenodo.org/record/4428917>

<sup>2</sup> <https://www.masoutis.gr/>

<sup>3</sup> <https://cloud.google.com/vision>

description of higher descriptive power (compared to PD), denoted hereafter as EPD. Consequently the EPD for Product SKU 18283, would be *Coffee Filtered Jacobs Vanilla Flavours 250 grams*.

## **1.2 Construction of the query image set**

The query images were captured under real supermarket conditions using a mobile phone camera and mimicking the scenario where visitors grab a product from the shelf with their hand to decide whether or not to put it in their cart. Google Pixel 2 XL with a camera resolution of 12.2 MP (3024 x 4032 resolution) was selected for acquiring the query images. Images were taken under different real-world lighting conditions and degree of background clutter (see Figure 1b), aiming to emulate in the best way the scenario of product selection from the visually impaired. Considering that the (majority of) visually impaired will not be able to identify a specific view of the product package (e.g. the frontal view), the set of query images was enriched with extra views for each product. The number of the captured images or views for each product was dictated by the type of the product package. The physical limitations imposed by slim and bag packages resulted in capturing only 2 images for products arising from these categories, while for the remainder of package types (e.g. box, can, bottle) 4 images were registered. An illustration of four different package types, arising from the query images and corresponding to the same SKUs illustrated in Figure 1a, is available in Figure 1b. The visual inspection of Figure 1 readily reveals the differences between the images captured in laboratory (i.e. candidate images) and natural (i.e. query images) environment. The set of query images was created by identifying a subset of the candidate SKUs, consisting of 104 randomly selected unique SKUs, resulting in a total of 373 query images. Finally, the annotation detail level for each query image was identical to the one for the candidate image set (i.e. BPC, SPC, BC, PD and EPD).