

Data Science Final Project Report

Andrés Ponce

Department of Computer Science

National Cheng Kung University

Tainan, Taiwan

andresponse@ismp.csie.ncku.edu.tw

Abstract—Many e-commerce platforms need to search for similar or identical products given some query image. Doing so can increase the platform’s ability to recommend interesting products or analyze purchasing trends across product categories. For the eBay eProduct Visual Search Challenge, participants take a set of query images and search a large index set for matching products. We first train a model to recognize the hierarchical structure of the different image categories. Then, we use our model’s output to produce hashes of the index images and query images, and locate identical products by comparing these hashes. This paper describes our method, experiments, and results used in this competition.

- [2] Jiangbo Yuan, An-Ti Chiang, Wen Tang, et al. *eProduct: A Million-Scale Visual Search Benchmark to Address Product Recognition Challenges*. 2021. DOI: 10.48550/ARXIV.2107.05856. URL: <https://arxiv.org/abs/2107.05856>.

I. INTRODUCTION

E-commerce platforms continue to grow and play a large role in consumer’s shopping behavior. Especially with the pandemic, more people relied on such platforms for many of their purchases [1].

E-commerce platforms that allow users to sell their own products especially require finding images of identical products. When a user searches for a product, he or she expects the results to contain images of the same product. On sites like eBay, identifying identical products can be very useful when aggregating sales of different listings of the same product. Not only do e-commerce platforms rely on such visual search, but also visual search engines such as Google Images, where the user can use an image as a query instead of a search term.

The eBay eProduct Visual Recognition Challenge [2] consists of finding images of the same product from a large index set of images. This challenge is one of fine-grained visual classification, since we are trying to find images of *the same* product. Similar products can differ by very small details, increasing the difficulty of the task. Likewise, identical products might differ based on lighting conditions or other yet we our model should still identify identical products despite these factors. pwe are trying to find images of *the same* product. Similar products can differ by very small details, increasing the difficulty of the task. Likewise, identical products might differ based on lighting conditions or other factors, yet our model should still identify identical products despite these factors.

REFERENCES

- [1] Petra Jílková and Petra Králová. “Digital consumer behaviour and ecommerce trends during the COVID-19 crisis”. In: *International Advances in Economic Research* 27.1 (2021), pp. 83–85.