SCALING CROSS-DOMAIN CONTENT-BASED IMAGE RETRIEVAL FOR E-COMMERCE SNAP AND SEARCH APPLICATION



Isaac Chung, Minh Tran, Eran Nussinovitch

BACKGROUND

E-commerce companies are increasingly adopting Computer Vision based technologies to offer a more appealing shopping experience and reduce shopper bounce rate. Visual search recommendations in their mobile and web applications improve customer experience by reducing the time and effort needed for product searches. These efforts aim to grow the number of conversion opportunities, which would lead to increased revenue.



CHALLENGES

We achieve improvements using a combination of visual search and classification capabilities, working in 50 with Fortune tandem а e-commerce client. This allows users to use their own photos to search for products via the client's mobile app. This content-based image retrieval task presents a couple of problems.

SCALE

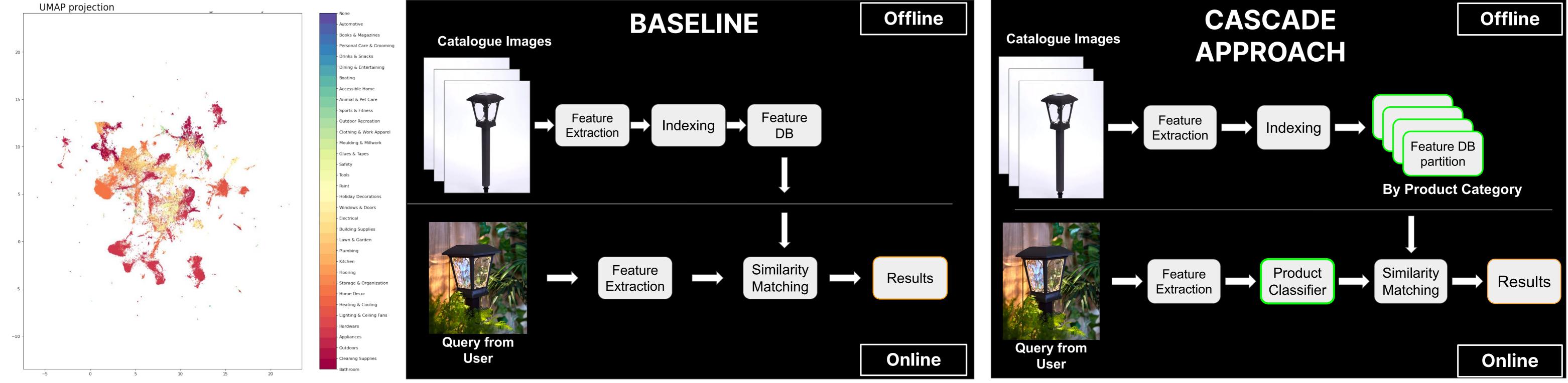
The search pool contains

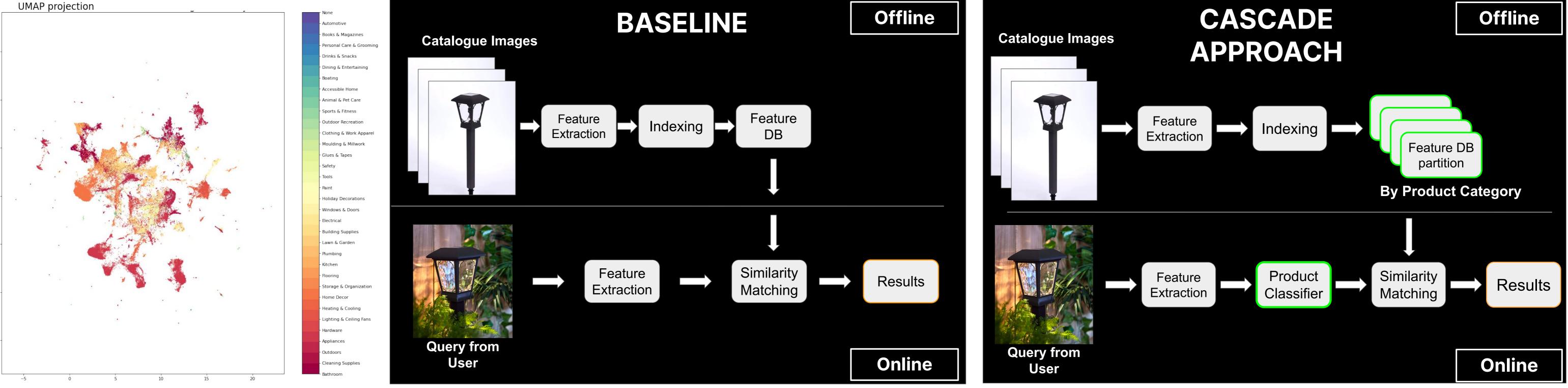
- 8M+ images
- 2M+ products
- 30+ Top Level Categories

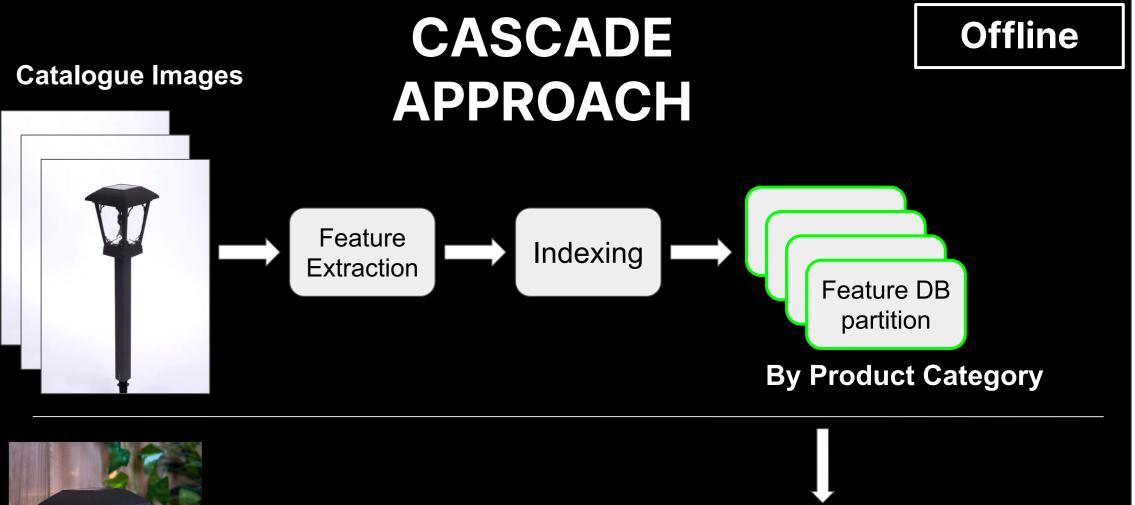
These categories were not designed with visual search in mind, but rather with traditional shopping.

CROSS DOMAIN



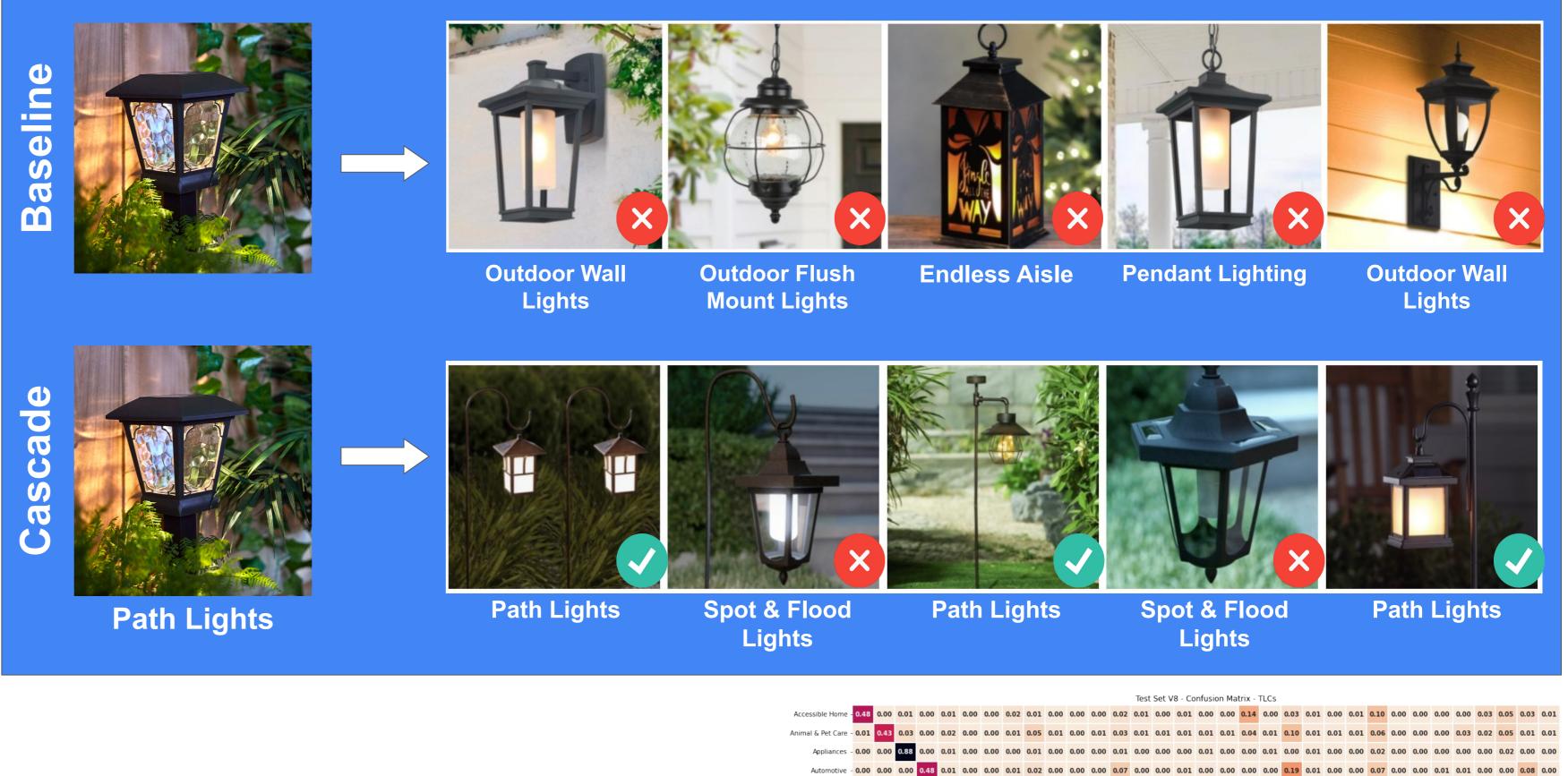






METHOD

Our approach addresses such problems by introducing an intermediate stage between the feature extractor and the approximate nearest neighbors (ANN) algorithm in the image retrieval system. This stage leverages the granular product hierarchy to perform a cascade-style search. The search pool is partitioned by TLC. A multi-class classifier, which is a convolutional neural network (CNN) trained on images in the same domain as the query data with labels as their respective TLCs, takes in query images as input and outputs TLC as predictions. Search is then performed in the search pool partition of the top prediction.



train

RESULTS

We are able to improve image retrieval metrics by an average of 69.7% compared to our baseline method, while limiting the overall latency increase to only 13%.

FUTURE WORK

- 1. Matching in-the-wild vs. test data class distribution
- 2. Covering unexpected use cases a. Barcodes / QR
- b. Screenshots / catalog images c. Shelf labels / Packages etc 3. Why stop at top 1 prediction? 4. Multi-modal search



Boating - 0.00 0.07 0.07 0.07 0.00 0.07 0.00 0.07 0.00 0 Glues & Tapes - 0.00 0.01 0.00 0.01 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.03 0.02 0.21 0.01 0.00 0.00 0.00 0.00 0.01 0.02 0.01 0.28 0.00 0.02 0.01 0.00 0.00 0.05 0.00 1.43 0.03 0.03 0.04 0.01 0.03 0.03 0.00 0.00 0.02 0.00 0.00 0.01 0.00 0.01 0.04 0.03 0.03 0.01 0.07 0.01 0.00 0.01 0.32 0.01 0.04 0.07 0.02 0.04 0.00 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.01 0.00 0.01 0.00 0.00 0.00 0.00

Corifoi

The deep learning platform for unstructured data