Attention-driven Unsupervised Image Retrieval for Beauty Products with Visual and Textual Clues

Jingwen Hou, Sijie Ji, Annan Wang

School of Computer Science and Engineering
Nanyang Technological University

Table of Contents

Background
- Beauty and Personal Care Product Retrieval
- Previous Approaches

Motivation

Proposed method
- Method Overview
- Initial stage: searching with visual clues
- Textual Index
- Refinement Stage: Refine the Initial Result with Textual Index

Experiment results

Table of Contents

Background
- Beauty and Personal Care Product Retrieval
- Previous Approaches

Motivation

Proposed method
- Method Overview
- Initial stage: searching with visual clues
- Textual Index
- Refinement Stage: Refine the Initial Result with Textual Index

Experiment results

Beauty and Personal Care Product Retrieval (BPCR)

Query image

Database images of the same item with label

Volumising Conditioner (For Fine or Flat Hair) 500ml/17.7oz

볼류마이징 컨디셔너 500ml

Aesop Volumising Conditioner

Peter Thomas Roth Camu Camu Power Cx30 Vitamin C Brightening Sleeping Mask 100ml/3. 4oz

Camu Camu Power C x 30™ Vitamin C Brightening Sleeping Mask

Peter Thomas Roth Camu Camu Power Sleeping Mask, 3.4 Fluid Ounce

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Previous Approaches

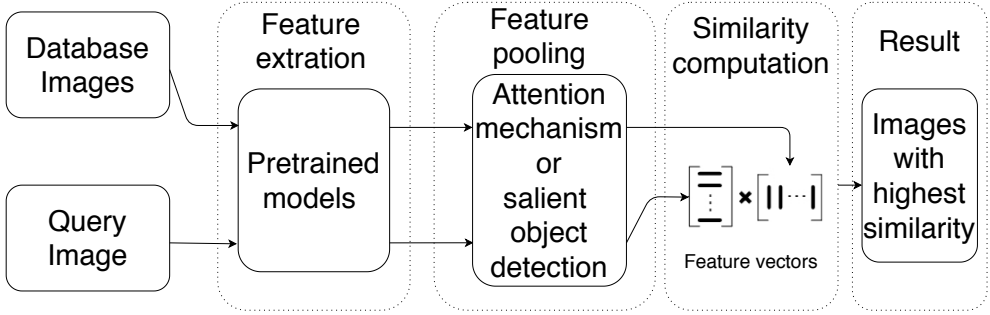

Figure: Common methods of beauty product retrieval problem.

Table of Contents

Background
- Beauty and Personal Care Product Retrieval
- Previous Approaches

Motivation

Proposed method
- Method Overview
- Initial stage: searching with visual clues
- Textual Index
- Refinement Stage: Refine the Initial Result with Textual Index

Experiment results

Subtle Visual Difference But Different Items

Some visual differences are subtle but crucial.

Figure: Images of lipsticks with similar packaging, the difference among them is hard to be noticed with non-specialized CNN.

Problem with CNN: Oversight of Subtle Visual Difference

Figure: Top 1 solution of 2019 challenge. The first 3 matched examples are very accurate. However, obviously false-positive matchings appear in the last 4 examples, implicating the inability of pretrained CNN to capture subtle visual differences.

Textual Information: Key to Tell the Difference

Visual differences neglected by CNN can be captured in product descriptions.

Christian Dior Rouge Dior Couture Colour Voluptuous Care - # 169 Grege 1947 3.5g/0.12oz

C2P Professional Make-Up Lipstick (4.5 g, 28)

Christian Dior Rouge Dior Couture Colour Voluptuous Care Lipstick for Women, No. 475 Rose Caprice, 0.12 Ounce

Figure: The aforementioned images of lipsticks with their labels. There are abundant information hard to be captured visually, such as brand in red or the color No. in blue.

Table of Contents

Background
- Beauty and Personal Care Product Retrieval
- Previous Approaches

Motivation

Proposed method
- Method Overview
- Initial stage: searching with visual clues
- Textual Index
- Refinement Stage: Refine the Initial Result with Textual Index

Experiment results

Method Overview

Figure: Search results before (top) and after (bottom) the proposed refinement. The proposed refinement strategy runs a second search within the examples with product descriptions similar to the top 3 matched examples of the first search result and replaces the last 4 matched examples of the first search result with the top 4 matched examples of the second search result.

Initial Stage: Searching with Visual Clues

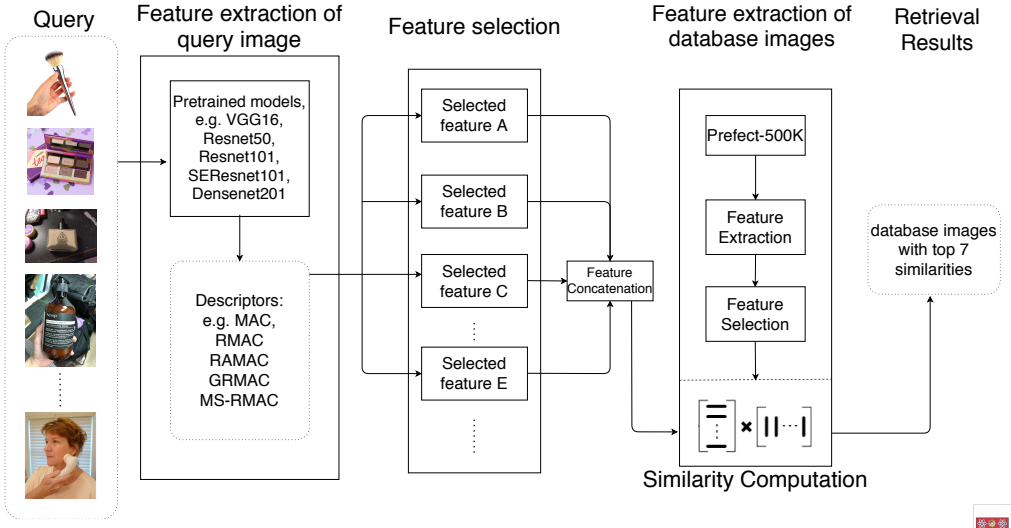

Figure: Overview of visual searching method (Yu et al. 2019).

Textual Index

The textual index (TI) is constructed by first vectoring each image label:

$$w_{t,d} = TF_{t,d} \times IDF_t, \qquad (1)$$

Then,for each document, find its similar documents by cosine similarity $s_{textual}$ across the collection:

$$s_{textual}(\mathbf{w}_{query}, \mathbf{w}_{data}) = \frac{\mathbf{w}_{query} \cdot \mathbf{w}_{data}}{\|\mathbf{w}_{query}\|_2 \|\mathbf{w}_{data}\|_2}. \qquad (2)$$

The Textual Index is put into a hash map, whose keys are image IDs and values are lists of image IDs of similar images.

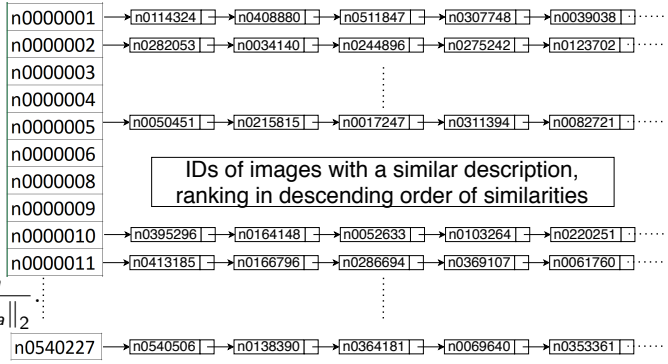

Figure: Part of the hash map of Textual index.

Refinement Stage: Refine the Initial Result with Textual Index

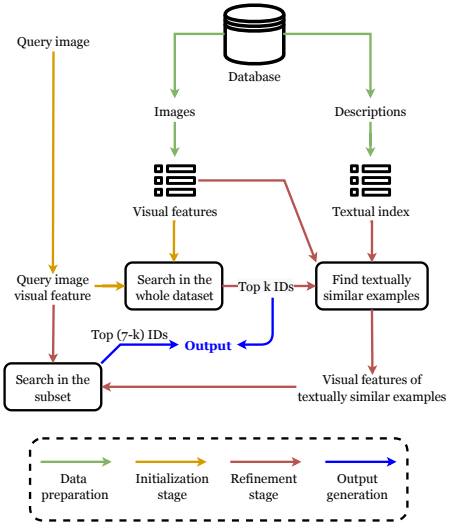

Figure: Overview of the proposed method.

Algorithm 1: Two-stage searching

Input : Query visual feature $\mathbf{v}_{query}$, textual index $\mathcal{D}(\cdot)$, database features $\mathbf{V}_{\mathcal{A}}$, parameter k

Output: Matched IDs $\mathcal{I}_{output} = \{jm\}_{m=1}^{7}$

$\mathcal{I}_{query,\mathcal{A}} \leftarrow argsortSimilaritiesDscd(\mathbf{v}_{query}, \mathbf{V}_{\mathcal{A}})$;

Initialize an empty list $\mathcal{I}_{subset}$;

for $n \leftarrow 1$ **to** k **do**

  Add all elements of $\mathcal{D}(\mathcal{I}^n_{query,\mathcal{A}})$ to $\mathcal{I}_{subset}$;

end

$\mathbf{V}_{subset} \leftarrow findFeaturesByIDs(\mathcal{I}_{subset}, \mathbf{V}_{\mathcal{A}})$;

$\mathcal{I}_{query,subset} \leftarrow argsortSimilaritiesDscd(\mathbf{v}_{query}, \mathbf{V}_{subset})$;

Initialize an empty list $\mathcal{I}_{output}$;

for $n \leftarrow 1$ **to** 7 **do**

  if $n <= k$ **then**

    Add $\mathcal{I}^n_{query,\mathcal{A}}$ to $\mathcal{I}_{output}$

  else

    Add $\mathcal{I}^{n-k}_{query,subset}$ to $\mathcal{I}_{output}$

  end

end

Table of Contents

Background
- Beauty and Personal Care Product Retrieval
- Previous Approaches

Motivation

Proposed method
- Method Overview
- Initial stage: searching with visual clues
- Textual Index
- Refinement Stage: Refine the Initial Result with Textual Index

Experiment results

Ablation Studies

Method	mAP@7	Improved %	Impaired %
Baseline	0.396944	-	-
Refined, k=6	0.397659	3	2
Refined, k=5	0.400262	6	2
Refined, k=4	0.405986	9	3
Refined, k=3	**0.407997**	**11**	**4**
Refined, k=2	0.402885	10	8
Refined, k=1	0.397293	7	6

Table: Results of ablation study on the validation set.

Qualitative Results

Figure: Initial results (left) and refined results (right) when k=3.

Thank You!

References I

Yu, Jun et al. (2019). "Beauty Product Retrieval Based on Regional Maximum Activation of Convolutions with Generalized Attention". In: *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2553–2557.