Fine-Grained Sketch-Based Image Retrieval: The Role of Part-Aware Attributes

KE LI

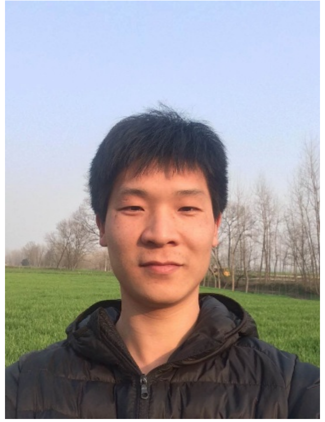
SketchX Lab@QMUL

PRIS@BUPT

北京邮电大学
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

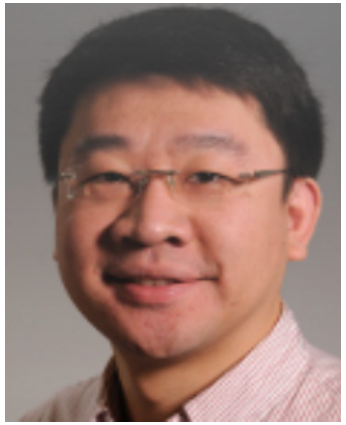
Queen Mary
University of London

Authors



Ke Li

Kaiyue Pang



Yi-Zhe Song

Timothy Hospedales

Honggang Zhang

Yichuan Hu

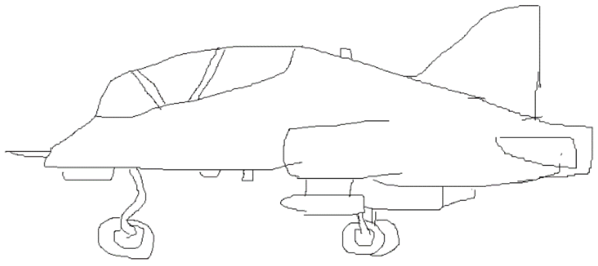
北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

Queen Mary
University of London

Motivation

- Sketches are intuitive and descriptive, which offers a more natural way to provide detailed visual cues than pure text.

Express in sketch



Express in pure text

airplane: military-plane, propeller-plane, on ground, small, single wing, tail-has-engine, two-wheel-1-axel…

Motivation

- Fine-grained sketch-based image retrieval (SBIR) is most likely to underpin practical commercial adoption of SBIR technology.

- There is lack of a purpose built fine-grained SBIR dataset to drive systematic research.

Why difficult?

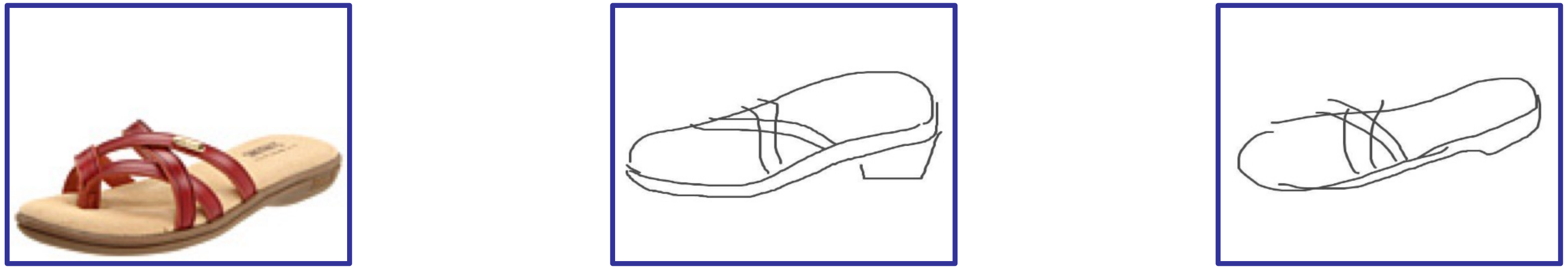
- Free-hand sketches are highly abstract and iconic.

Why difficult?

- Fine-grained correspondence between sketches and images is difficult to establish especially given the abstract and cross-domain nature of the problem.

Boot

Sandal



Our Contribution

- We propose a fine-grained SBIR shoe dataset with free-hand human sketches and photos, as well as fine-grained attribute annotations.
- We propose a part-aware paradigm that allows fine-grained attribute detection.
- We propose a synergistic low-level + mid-level + high-level feature representation that proves to crucial to improve the performance of fine-grained SBIR.

Our Proposed Dataset

- Fine-grained sketch image pair:

Our Proposed Dataset

• Various drawing styles

Image | Free-hand sketch

Drawer 1 Drawer 2 Drawer 3

Image | Free-hand sketch

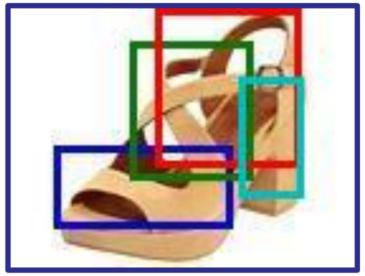
Drawer 1 Drawer 2 Drawer 3

Our Proposed Dataset

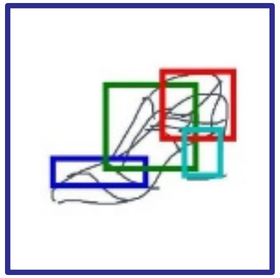
- Dataset stats

(1) 304 images and 912 free-hand human sketches with each image having three fine-grained correspondings.

(2) Comprehensive part attribute, bounding box annotations



{1,0,0,1,1 …}



{1,0,0,1,1 …}

Why Part-aware?

- Problem
  - Attribute co-occurrence patterns may differ from what is observed in training, given part-based attributes.

Training case

Test failure case

Attribute that should be learnt

Co-occurance pattern that actually learnt

Why Part-aware?

- Possible Solutions

—Attribute decorrelation by inducing

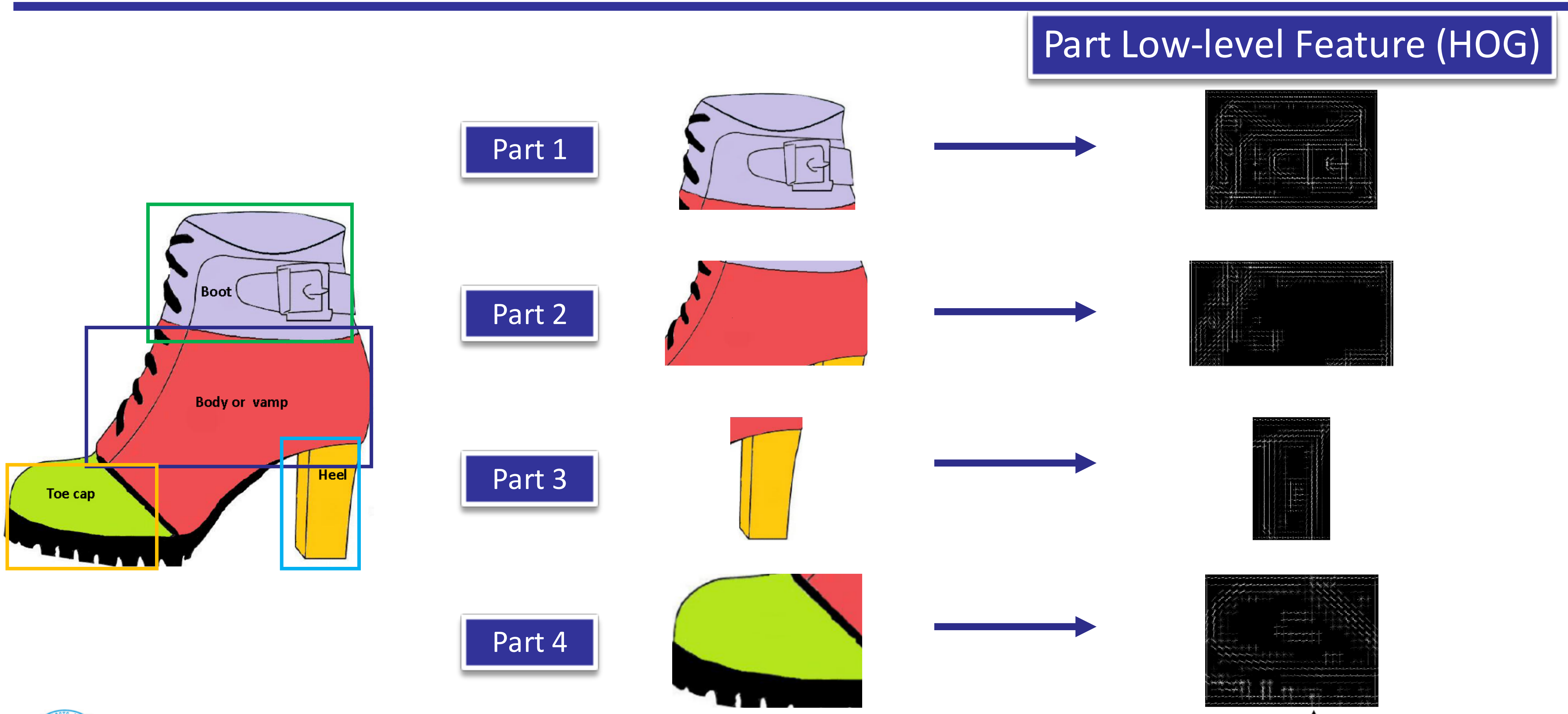
(1) in-group feature sharing

(2) between-group competition for features. (Structural sparsity problem)

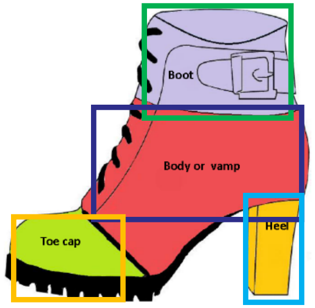
—Part-aware attributes by applying

Deformable Part-based model to perform attribute detection within each part (group).

Generating a synergistic cross-domain representation: A part-aware approach

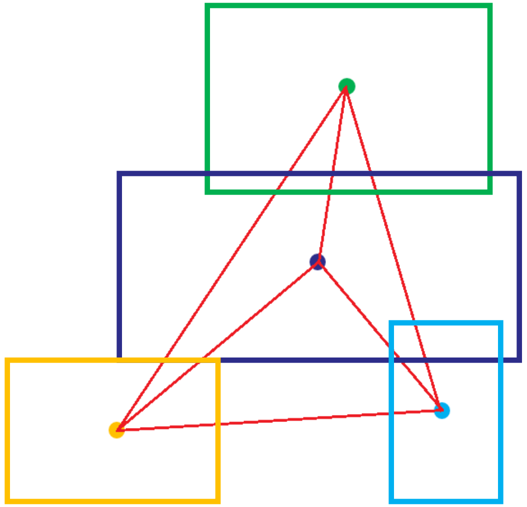


Generating a synergistic cross-domain representation: A part-aware approach



Part Structure

Part Attribute



Part 1 {low boot, middle boot, high boot } → {1,0,0}

Part 2 {ornament or brand on body side, ornament or shoelace on vamp } → {1,1}

Part 3 {round, toe-open } → {1,0}

Part 4 {low heel, High heel, pillar heel, cone heel, slender heel, thick heel } → {1,0,0,1,0,1}

Three View CCA

Experiments and Results

Whole-Image	WS-Decor[3]	WS-DPM	SS-Decor[3]	Ours	Ground-truth part
80.89%	78.66%	81.05%	81.72%	83.68%	86.19%

Table 1: Comparison with state-of-the-art attribute detection on image

Whole-Image	WS-Decor[3]	WS-DPM	SS-Decor[3]	Ours	Ground-truth part
74.91%	74.00%	73.12%	75.73%	75.89%	80.29%

Table 2: Comparison with state-of-the-art attribute detection on sketch

Experiments and Results

Part-Structure	Part-HOG	Part-HOG + Part-Structure + 2View-CCA	Part-Attribute
7.33%	17%	23.33%	20%
Part-Attribute + Part-Structure + 2View-CCA	**Part-HOG + Part-Attribute + 2View-CCA**	**Part-HOG + Part-Attribute + Part-Structure + 3View-CCA**	**Ground- truth**
26%	28.33%	33%	46.67%

Table 3: Performance of comparisons on fine-grained SBIR @ K = 5

Attribute detection corrected by our part-aware approach

Part aware		Whole
Has shoelace on vamp Has enclosed toe Has low boot Has low heel		Has nothing on vamp ✗ Has enclosed toe Has low boot Has high heel ✗

Part aware		Whole
Has nothing on vamp Has open toe Has low boot Has low heel		Has shoelace on vamp ✗ Has open toe Has low boot Has low heel

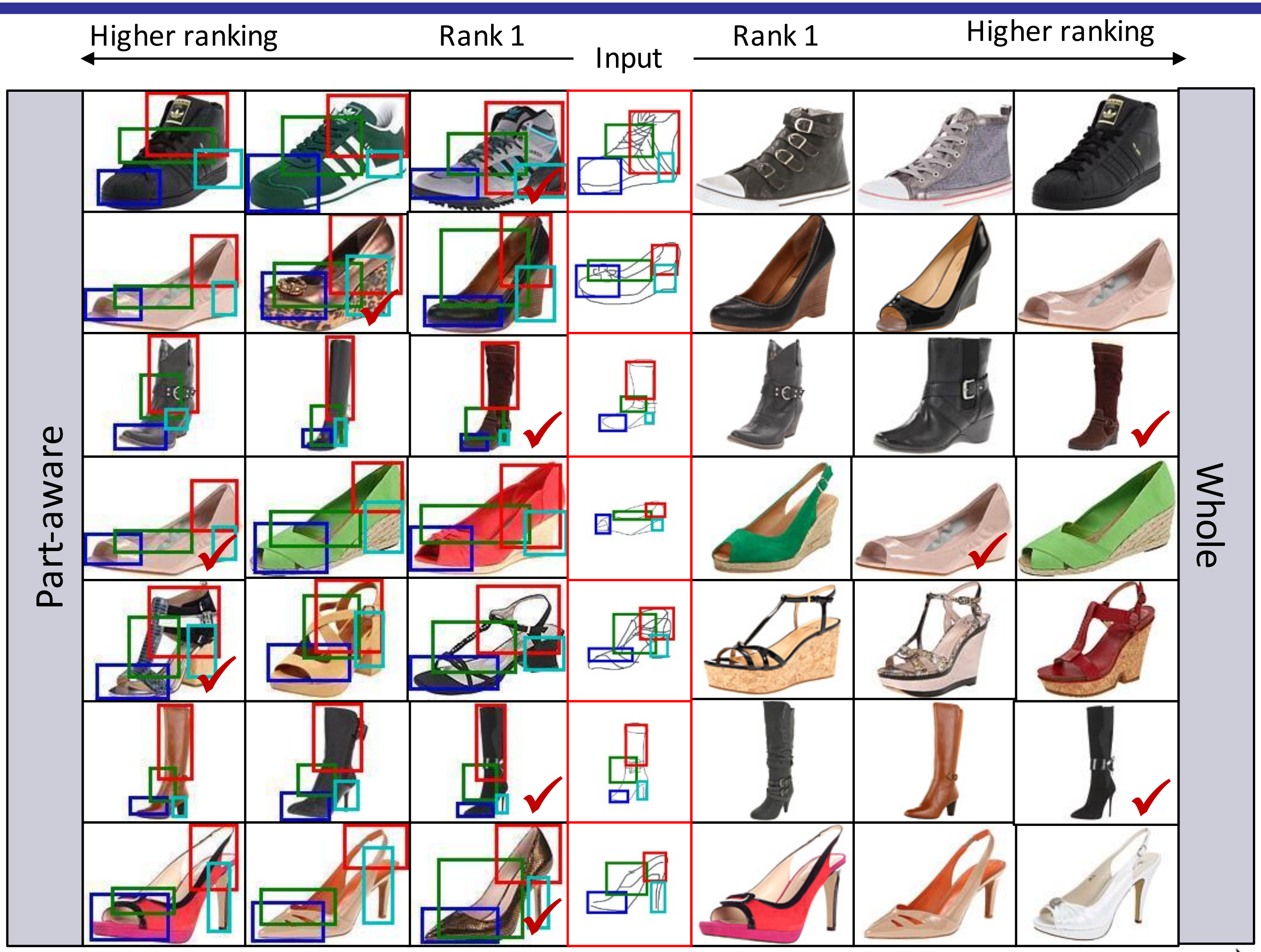
Part aware		Whole
Has nothing on vamp Has enclosed toe Has high boot Has high heel		Has shoelace on vamp ✗ Has enclosed toe Has high boot Has low heel ✗

Part aware		Whole
Has shoelace on vamp Has enclosed toe Has low boot Has low heel		Has shoelace on vamp Has enclosed toe Has middle boot ✗ Has low heel

Part aware		Whole
Has nothing on vamp Has open toe ✗ Has low boot Has high heel		Has nothing on vamp Has enclosed toe Has middle boot ✗ Has low heel ✗

Part aware		Whole
Has nothing on vamp Has enclosed toe Has high boot Has low heel		Has shoelace on vamp ✗ Has enclosed toe Has high boot Has low heel

Fine-grained SBIR with and without our proposed part-aware method



北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS
Queen Mary
University of London

References

- [1] Hossein Azizpour and Ivan Laptev. Object detection using strongly-supervised deformable part models. In ECCV, pages 836–849. 2012.
- [2] Yunchao Gong, Qifa Ke, Michael Isard, and Svetlana Lazebnik. A multi-view embedding space for modeling internet images, tags, and their semantics. IJCV, pages 210–233, 2014.
- [3] Dinesh Jayaraman, Fei Sha, and Kristen Grauman. Decorrelating semantic visual attributes by resisting the urge to share. In CVPR, pages 1629–1636, 2014.
- [4] Yi Li, Timothy M. Hospedales, Yi-Zhe Song, and Shaogang Gong. Fine-grained sketch-based image retrieval by matching deformable part models. In BMVC, 2014.
- [5] Shuxin Ouyang, Timothy Hospedales, Yi-Zhe Song, and Xueming Li. Cross-modal face matching: Beyond viewed sketches. In ACCV, pages 210–225. 2014.

Questions