Sketch Me That Shoe

Qian Yu et al.

CVPR 2016

presenter: Wei-Lin Hsiao

advisor: Kristen Grauman

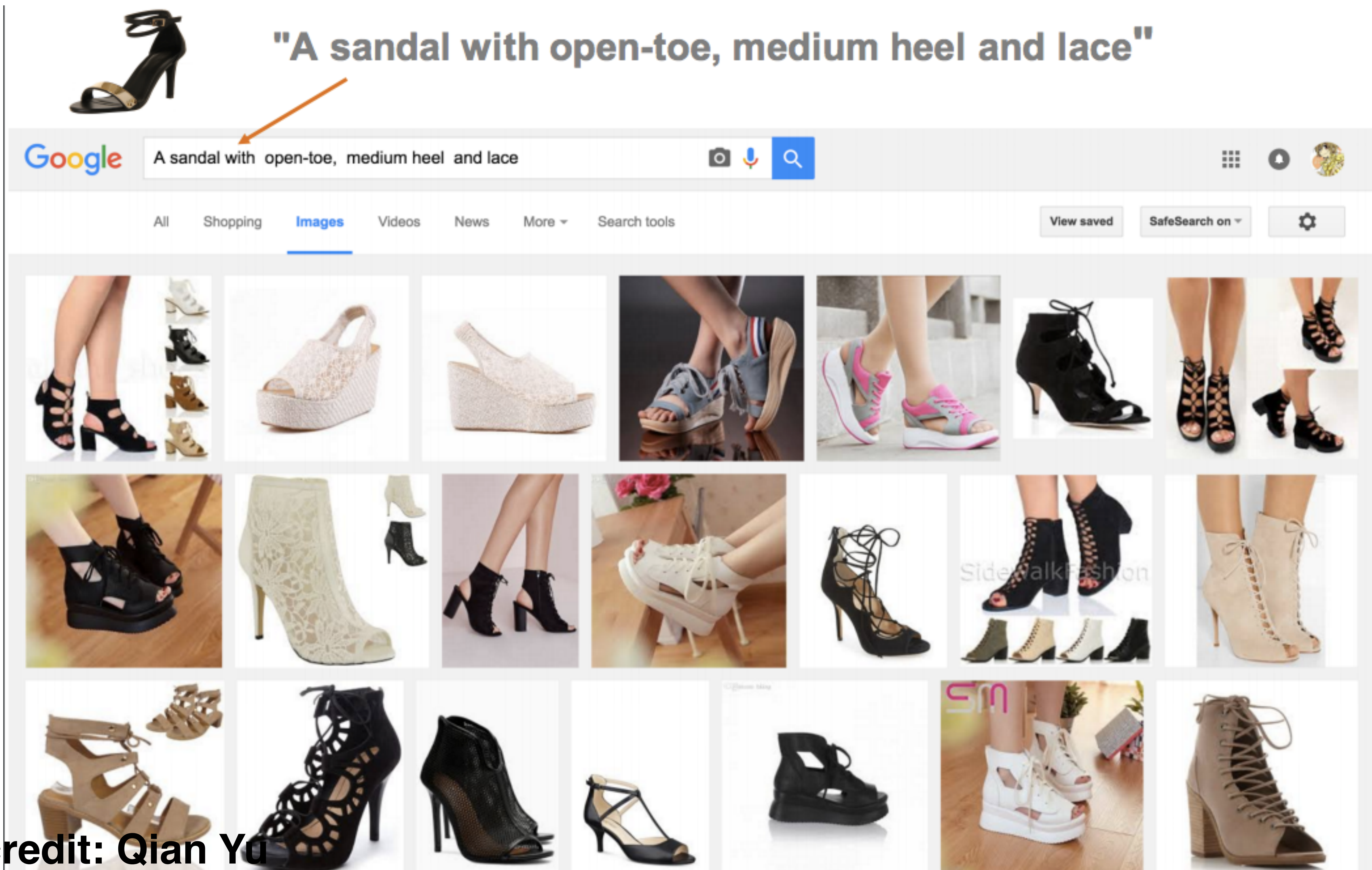
Traditional text-based image retrieval:

" A **sandal** with **open-toe**, medium **heel** and **lace** "

"A **ladies shoe**…"

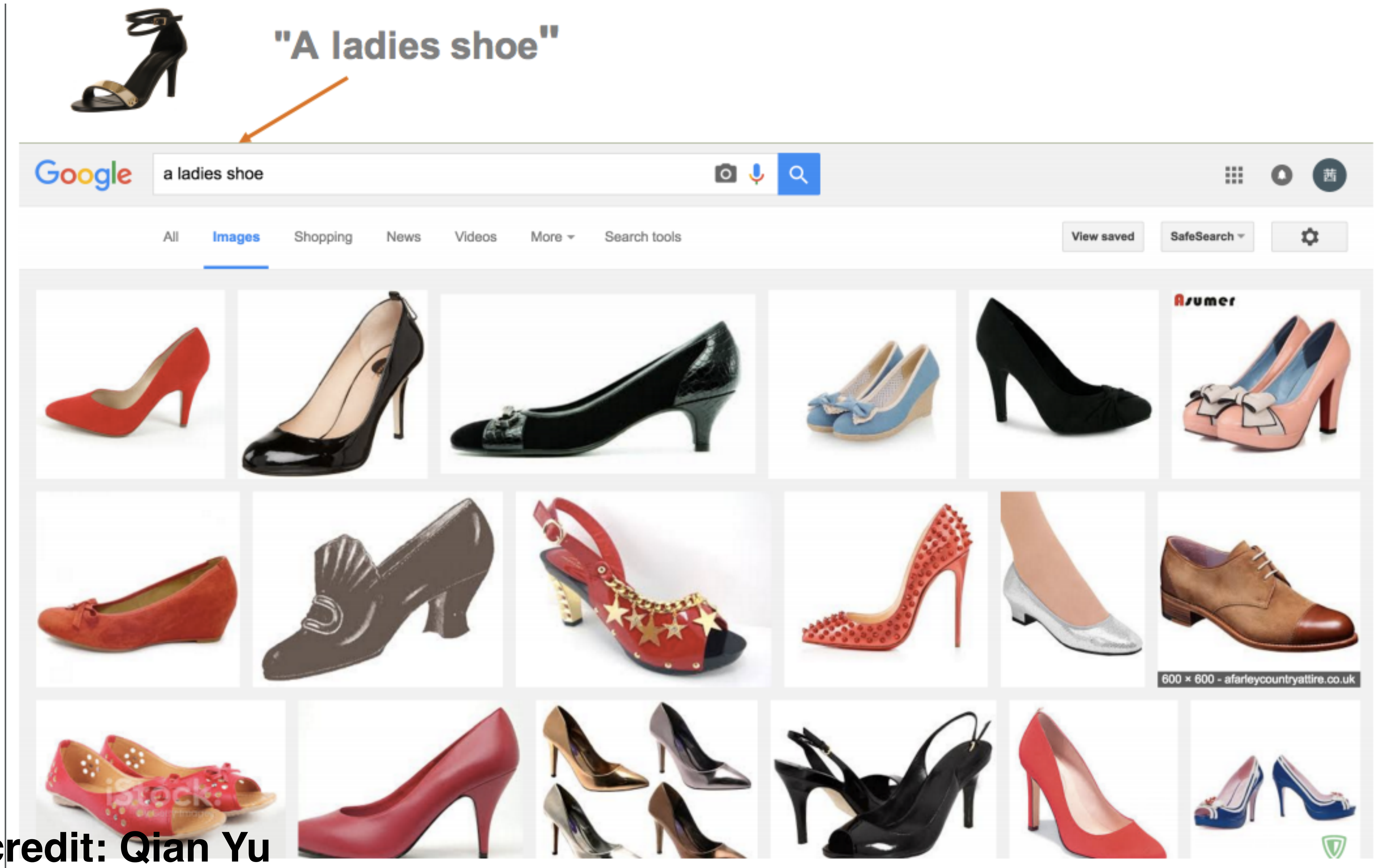
slide credit: Qian Yu

Image retrieval by **text** is challenging



slide credit: Qian Yu

Image retrieval by **text** is challenging



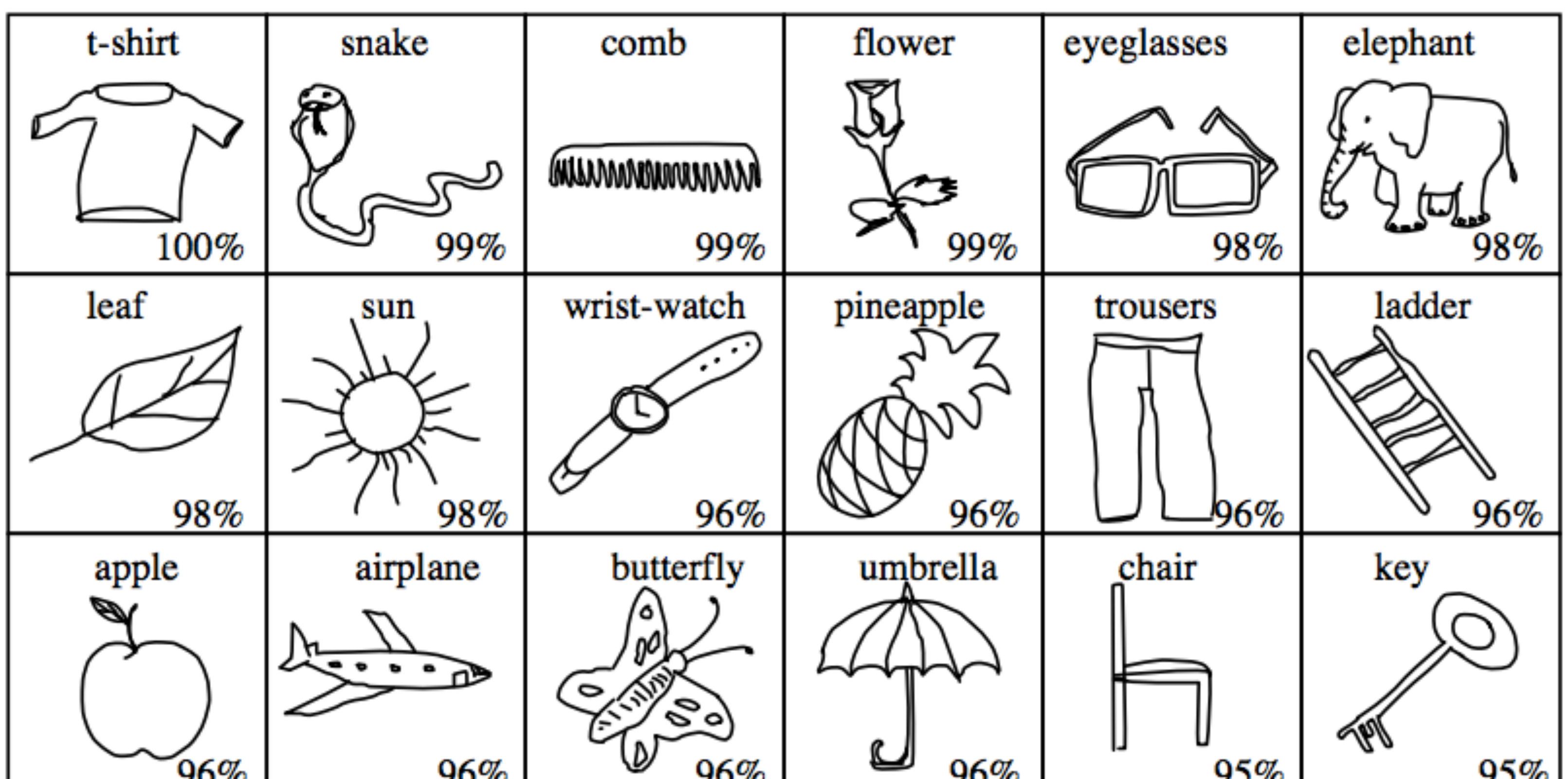
slide credit: Qian Yu

A sketch speaks for a hundred words

slide credit: Qian Yu

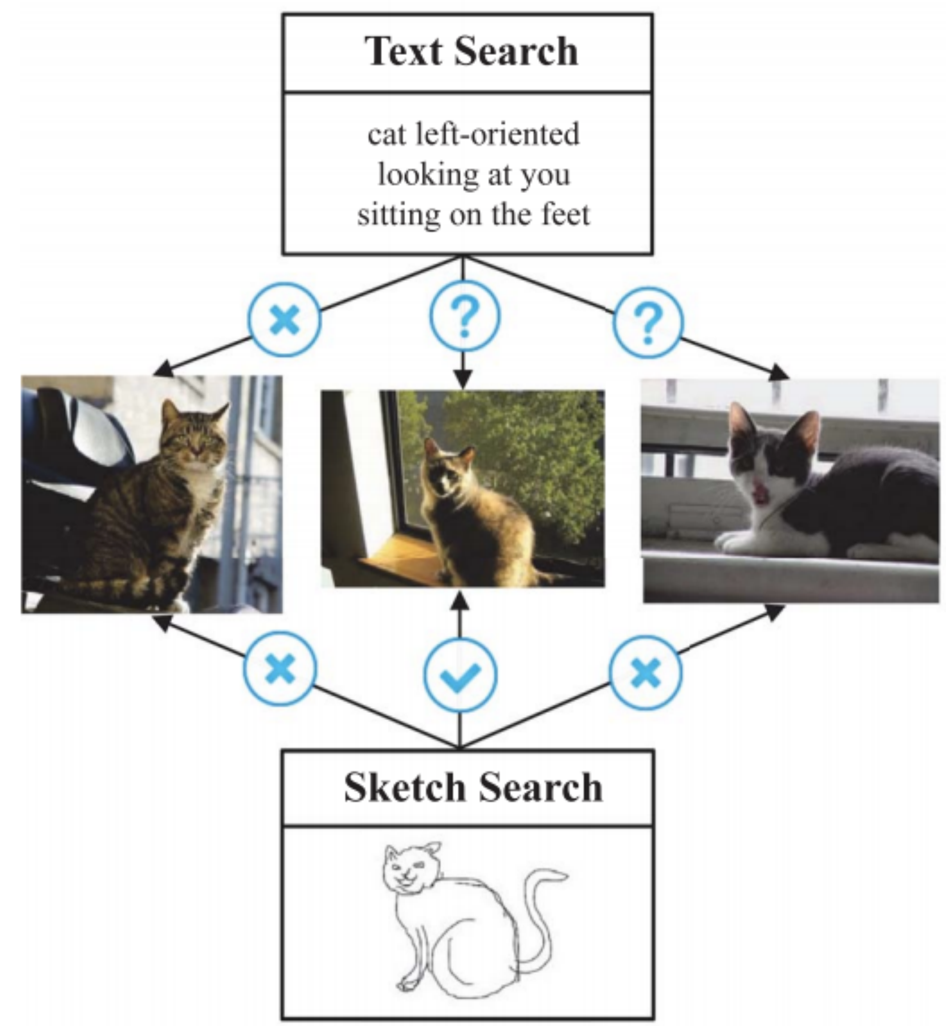
Sketch-based image retrieval (SBIR) — related work

- **Category-level** SBIR:
  - E. Mathis et al. TVCG 2011, E. Mathis et al. Computers & Graphics 2010, R. Hu ICIP 2010, Y. Cao, ACM 2010, ….



Sketch-based image retrieval (SBIR) — related work

- Fine-grained SBIR:
  - fine-grained in the way of **object configuration**
  - Y.Li, T. Hospedales, Y.-Z. Song, and S. Gong. fine-grained sketch-based image retrieval by matching deformable part models. In BMVC, 2014



Fine-grained instance-level sketch-based image retrieval (SBIR)

- Challenges

1. visual comparison in a **fine-grained**, **cross-domain** way

2. free-hand sketches are highly **abstract**

3. annotated cross-domain sketch-photo **datasets** are **scarce**

Main contribution

1. Introduce two new datasets

Main contribution

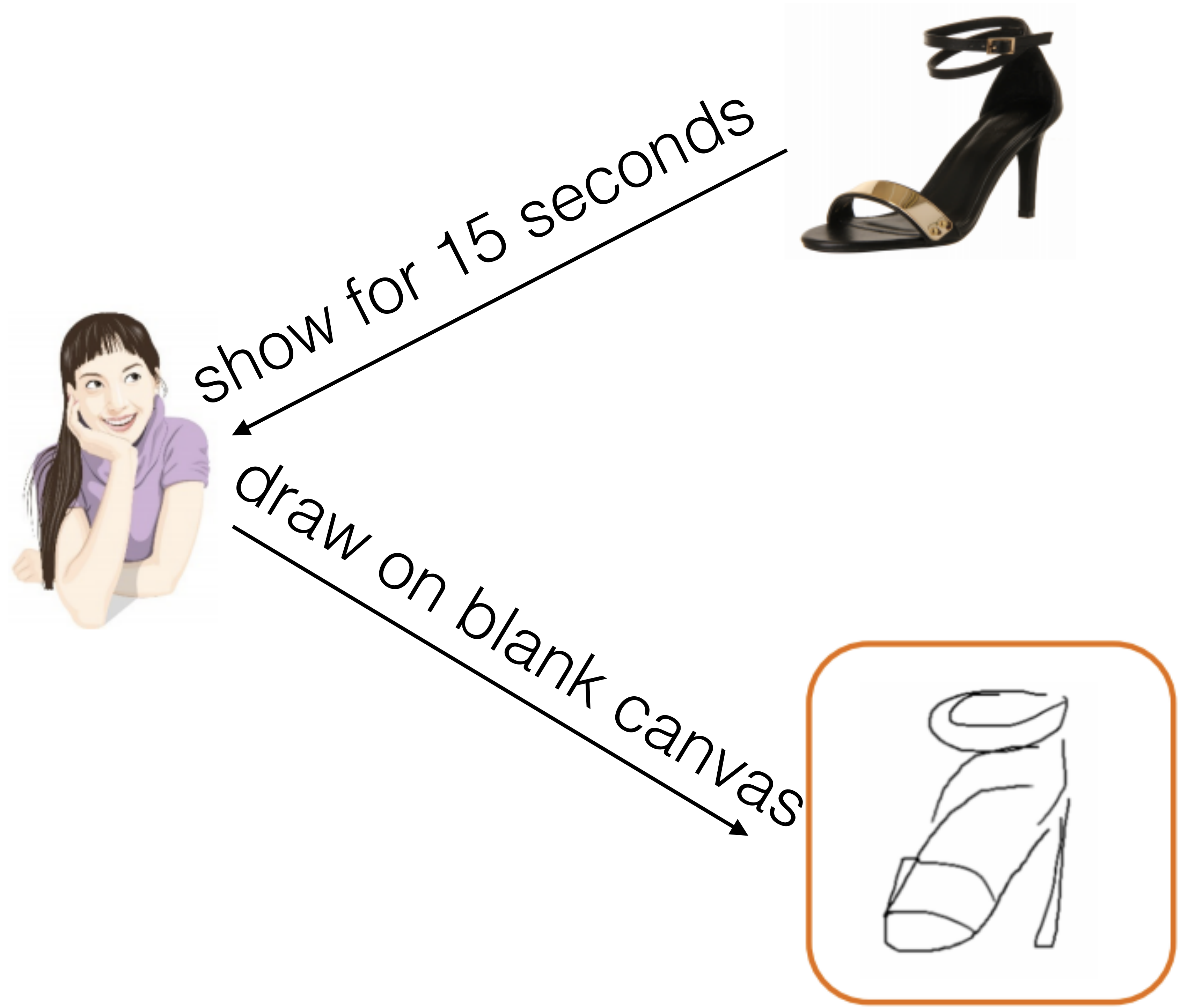
2. Overcome the requirements of extensive data and annotation by

- pre-training
- sketch-specific data augmentation

Data collection—photo images

- Shoe images
  - UT-Zap50K
  - 419 images, high-heel, ballerinas, formal, informal
- Chair images
  - IKEA, Amazon, Taobao
  - 297 images, office chairs, couches, kids chair, desk chairs…

Data collection—sketches



22 volunteers: none has any art training

Data annotation

- Train a **ranking model** instead of a verification model

- **Triplet ranking** instead of global ranking

    - given a sketch query, which of the two photos is more similar to it?

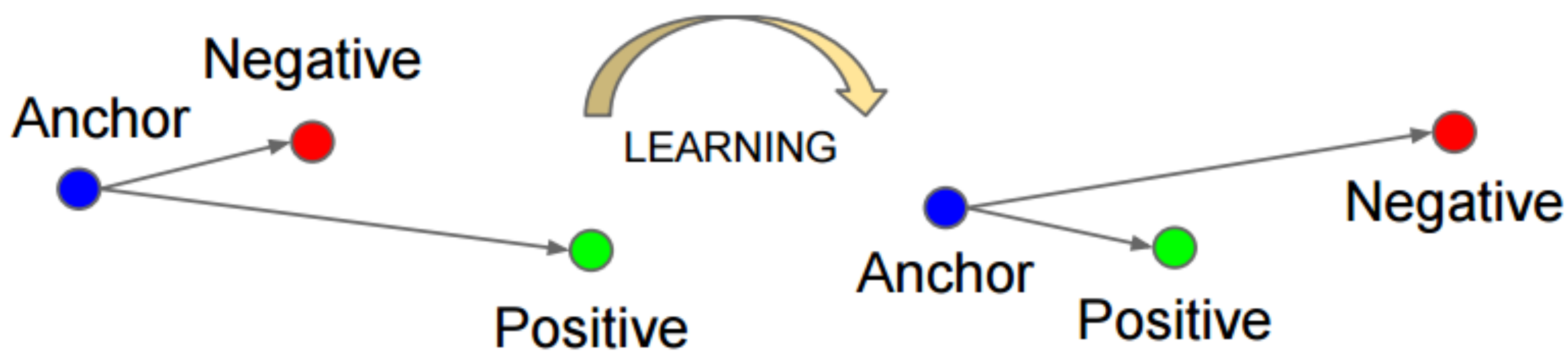
    - Question: How to select a subset of triplets to be annotated?

Data annotation

1. Attribute annotation:
    - Need to measure distance between a sketch and a photo
    - Based on: attribute vector + deep feature vector
2. Generating candidate photos for each sketch:
    - Top 10 closest photo images to the query sketch
3. Triplet annotation:
    - $C^{10}{}_{2}$ triplets for each sketch; 3 people annotated each triplet.
    - Majority voting to merge 3 annotations.

Objective function for triplet ranking

$$\min_{\theta} \sum_{t \in T} L_{\theta}(t) + \lambda R(\theta)$$

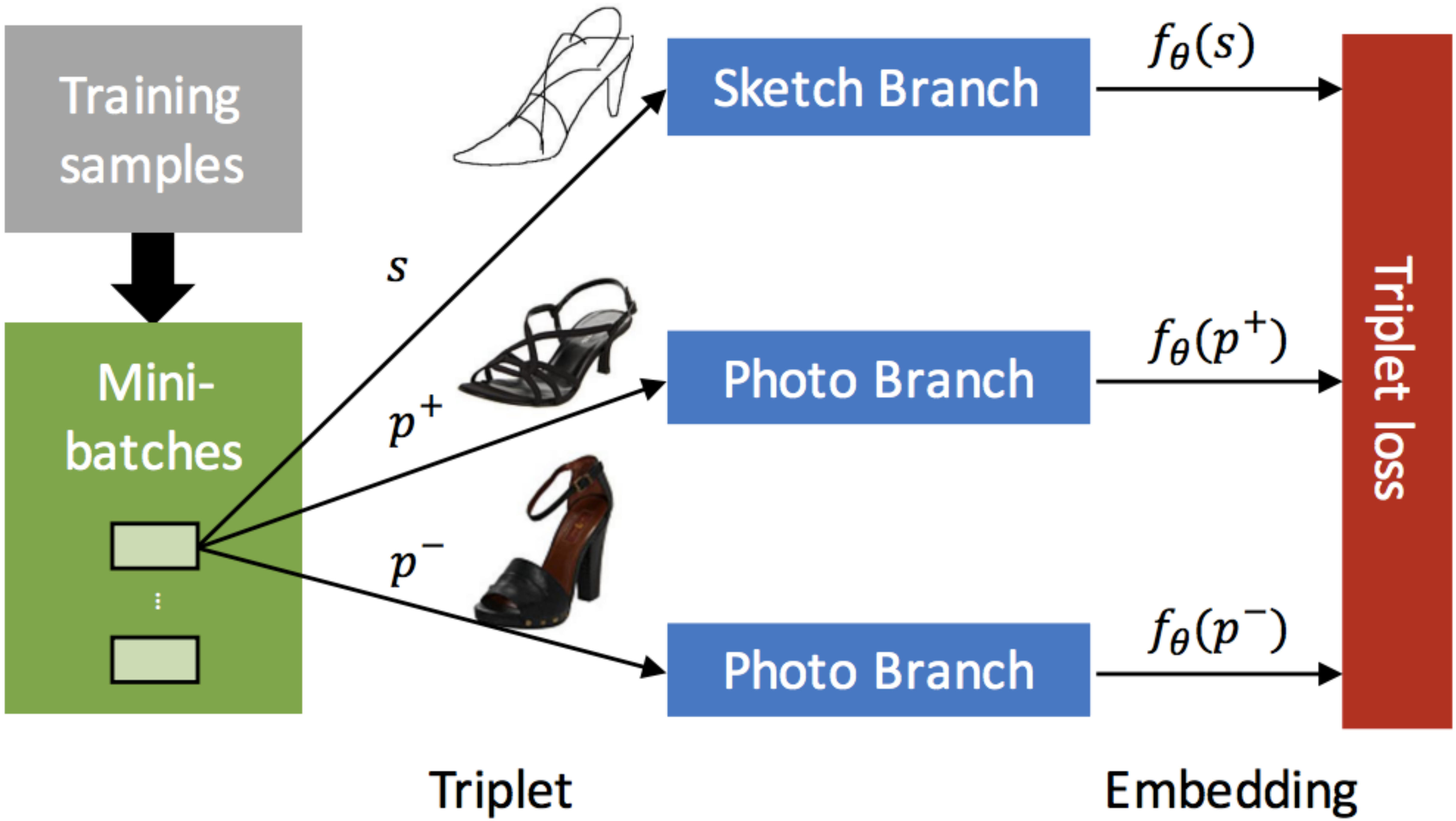


$$L_{\theta}(t) = \max(0, \Delta + D(f_{\theta}(s), f_{\theta}(p^{+})) - D(f_{\theta}(s), f_{\theta}(p^{-})))$$

distance between sketch and positive photo

distance between sketch and negative photo

Network architecture

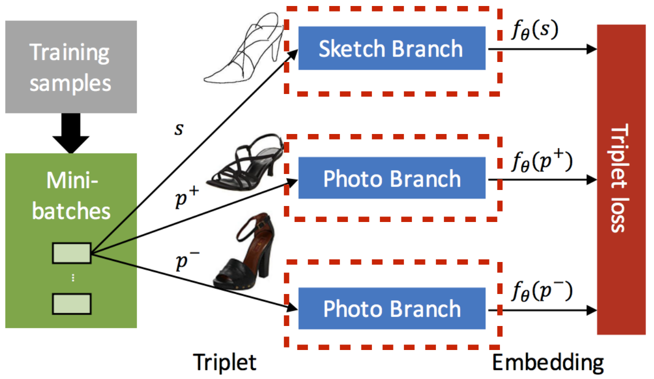


Pre-train/fine-tune

1. Generalize to both photos and sketches
2. Exploit auxiliary sketch/photo category-paired data to pre-train the ability to rank
3. Fine-tune on contributed shoe/chair dataset

Generalize to both photos and sketches— Step1,2

- Train a single Sketch-a-Net to recognize both photos and sketches



1. Photos:

    - Pre-train to classify 1000 categories of ImageNet-1K with **edge maps** extracted

2. Free-hand sketches:

    - Fine-tune to classify 250 categories of TU-Berlin

Sketch-a-Net that Beats Humans Q. Yu, Y. Yang, Y-Z. Song, T. Xiang and T. Hospedales(BMVC 2015)

Exploit auxiliary sketch/photo category-paired data—Step 3

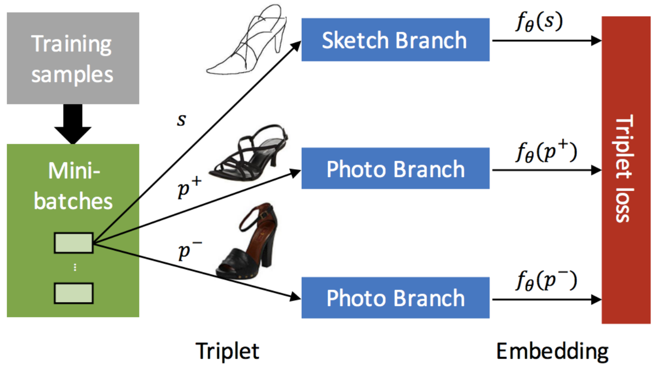
- Train sketch-photo <u>ranking network</u>:

  1. Initialize each branch network with the previous learned Sketch-a-Net

  2. Pre-train triplet ranking model using **category-level** annotation

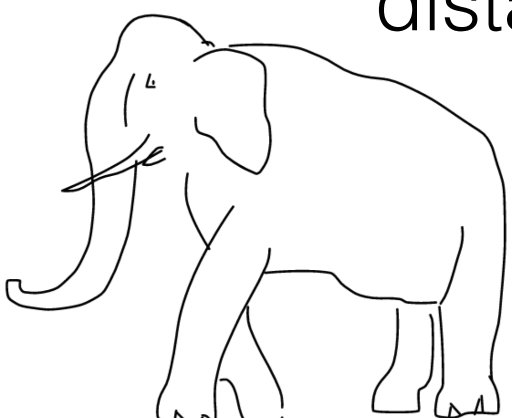
  - select 187 **categories** which exist in both TU-Berlin(sketch) and ImageNet(photo)

  - 8976 sketches, 19026 photos

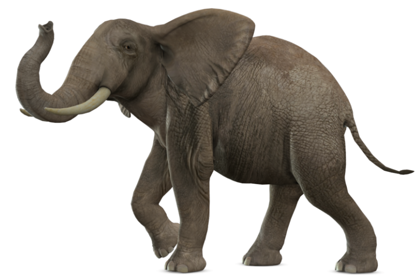


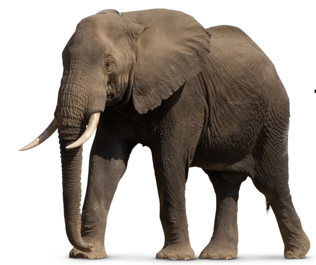
Exploit auxiliary sketch/photo category-paired data—Step 3

distance: Euclidean distance of Sketch-a-Net features



query sketch





top 20% most similar
same class

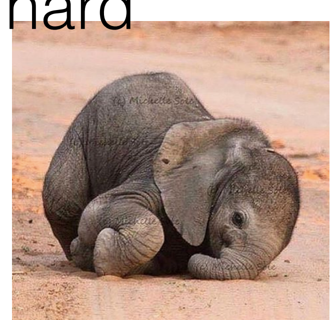
easy

random
different classes

out-of-class hard

distances smaller than positives
different classes

in-class hard

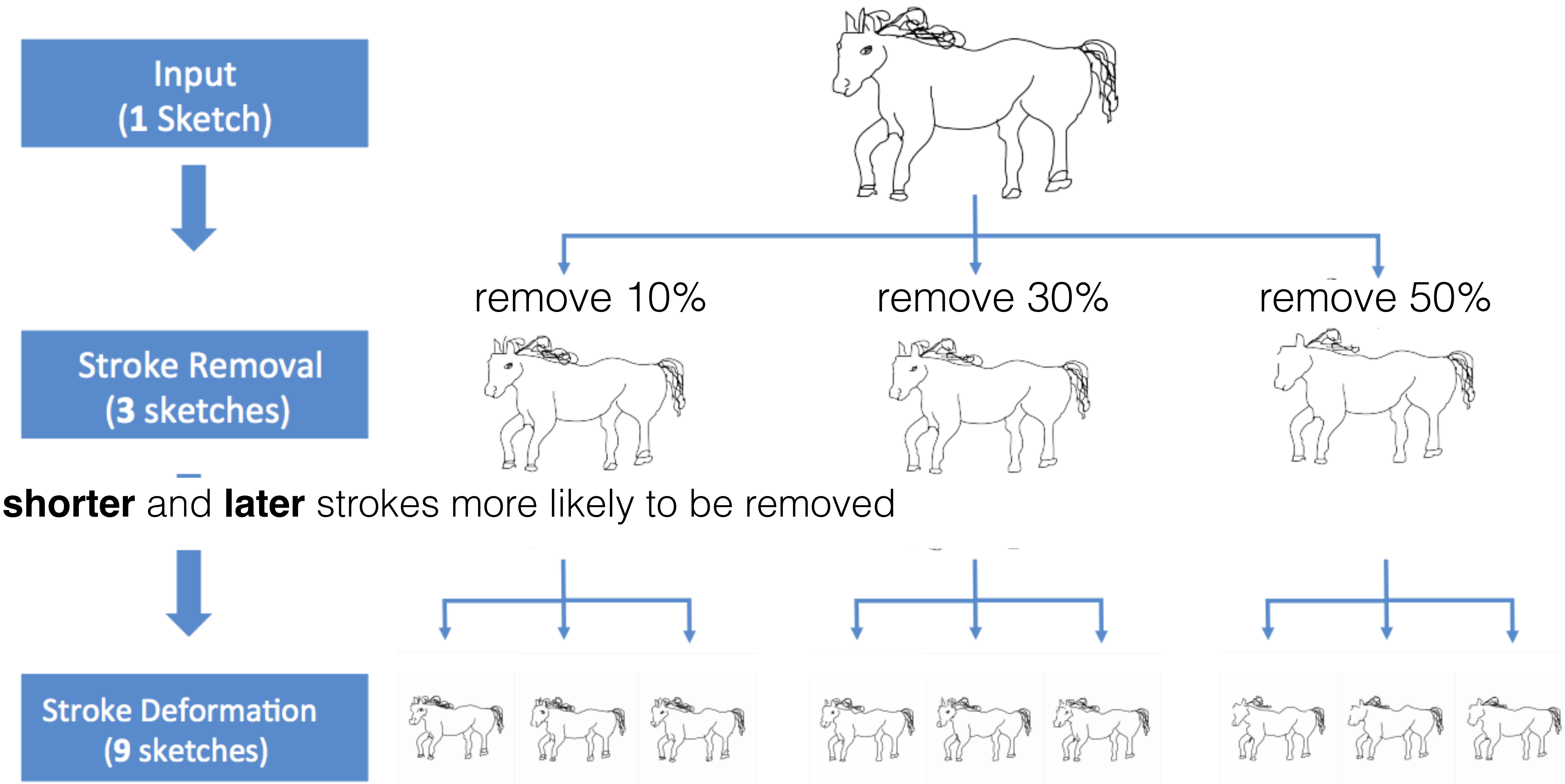


bottom 20% most similar
same class

Fine-tune on target scenario —Step 4

- Train sketch-photo ranking network:
  - Fine-tune on contributed shoe/chair dataset

Data augmentation



shorter and **smaller** curvature strokes are probabilistically deformed more

Experiments—fine-grained instance-level retrieval

- Evaluation metrics
    - **retrieval accuracy**: how quickly a model finds a specific item/image
    - **% correctly ranked triplets**: overall quality of a model's ranking list

Experiments—fine-grained instance-level retrieval

- Baselines
  - hand-crafted
    - HOG+BoW+RankSVM
    - Dense HOG+RankSVM
  - deep features
    - single Sketch-a-Net extracted feature
    - 3D shape: F.Wang, L.Kang, Y.Li, “Sketch-based 3d shape retrieval using convolutional neural networks”, CVPR 2015

Experimental result

random: 50%

Shoe Dataset	acc.@1	acc.@10	%corr.
BoW-HOG + rankSVM	17.39%	67.83%	62.82%
Dense-HOG + rankSVM	24.35%	65.22%	67.21%
ISN Deep + rankSVM	20.00%	62.61%	62.55%
3DS Deep + rankSVM	5.22%	21.74%	55.59%
Our model	**39.13%**	**87.83%**	**69.49%**
Chair Dataset	**acc.@1**	**acc.@10**	**%corr.**
BoW-HOG + rankSVM	28.87%	67.01%	61.56%
Dense-HOG + rankSVM	52.57%	93.81%	68.96%
ISN Deep + rankSVM	47.42%	82.47%	66.62%
3DS Deep + rankSVM	6.19%	26.80%	51.94%
Our model	**69.07%**	**97.94%**	**72.30%**

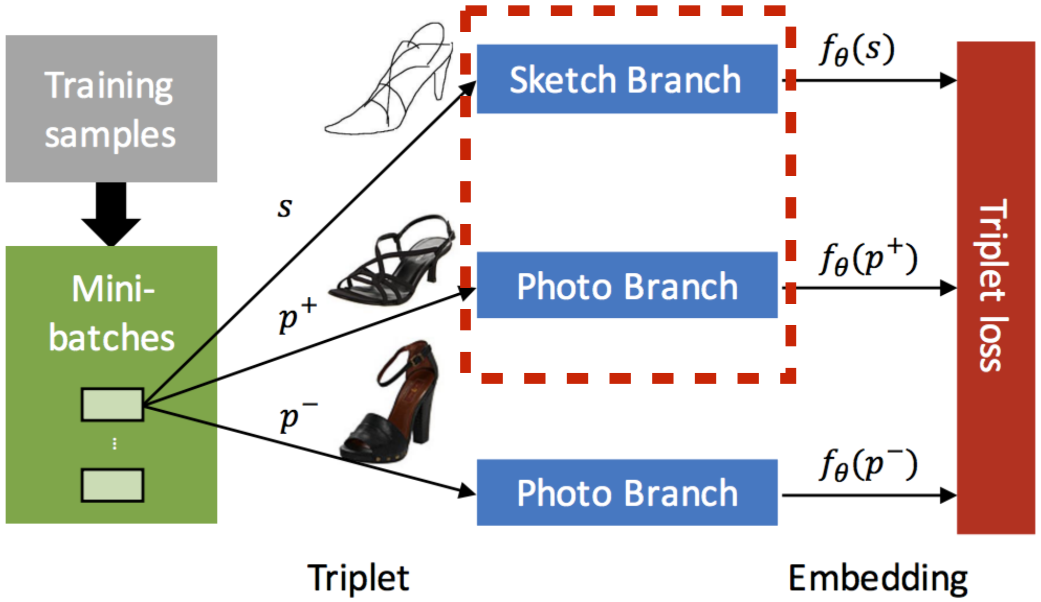
Experimental result

Contribution of different component

	acc.@1	acc.@10
without any pretaining Step 4 only	27.83%	78.26%
pre-train to generalize to sketch Step 2 + 4, no data aug	33.04%	81.74%
Step 2 + 4, with data aug	36.52%	84.35%
pre-train to generalize to photo Step 1 + 2 + 4, with data aug	38.26%	85.22%
Step 1-4, no data aug	37.39%	86.09%
Our full model	**39.13%**	**87.83%**

Siamese or heterogeneous?
Ranking or verification?

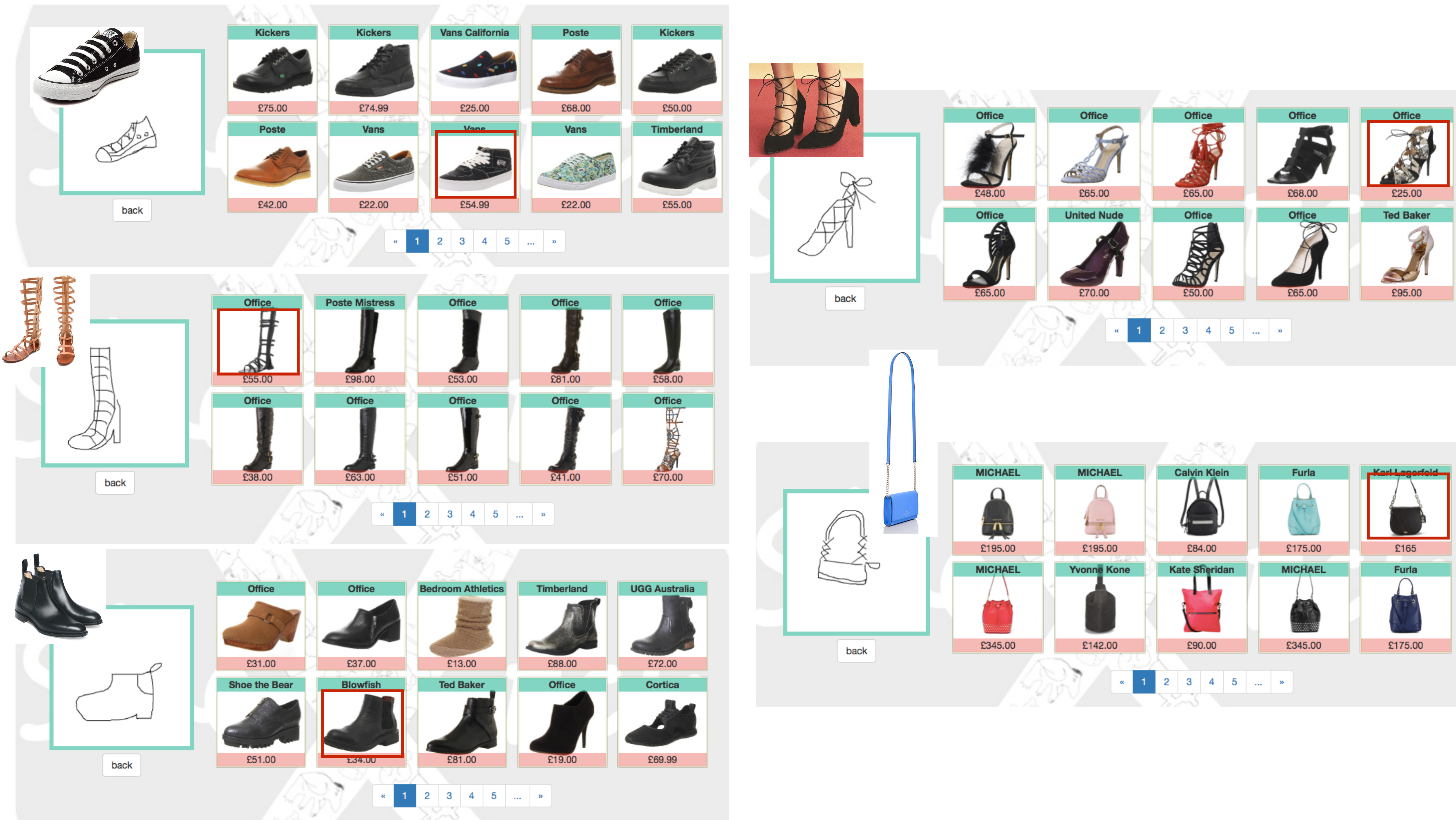
	acc.@1	acc.@10
Siamese verification	28.70%	78.26%
Hetero. **ranking**	$\boxed{21.74\%}$	$\boxed{68.70\%}$
Hetero. **verification**	16.52%	69.57%
Our full model siamese, ranking	**39.13%**	**87.83%**



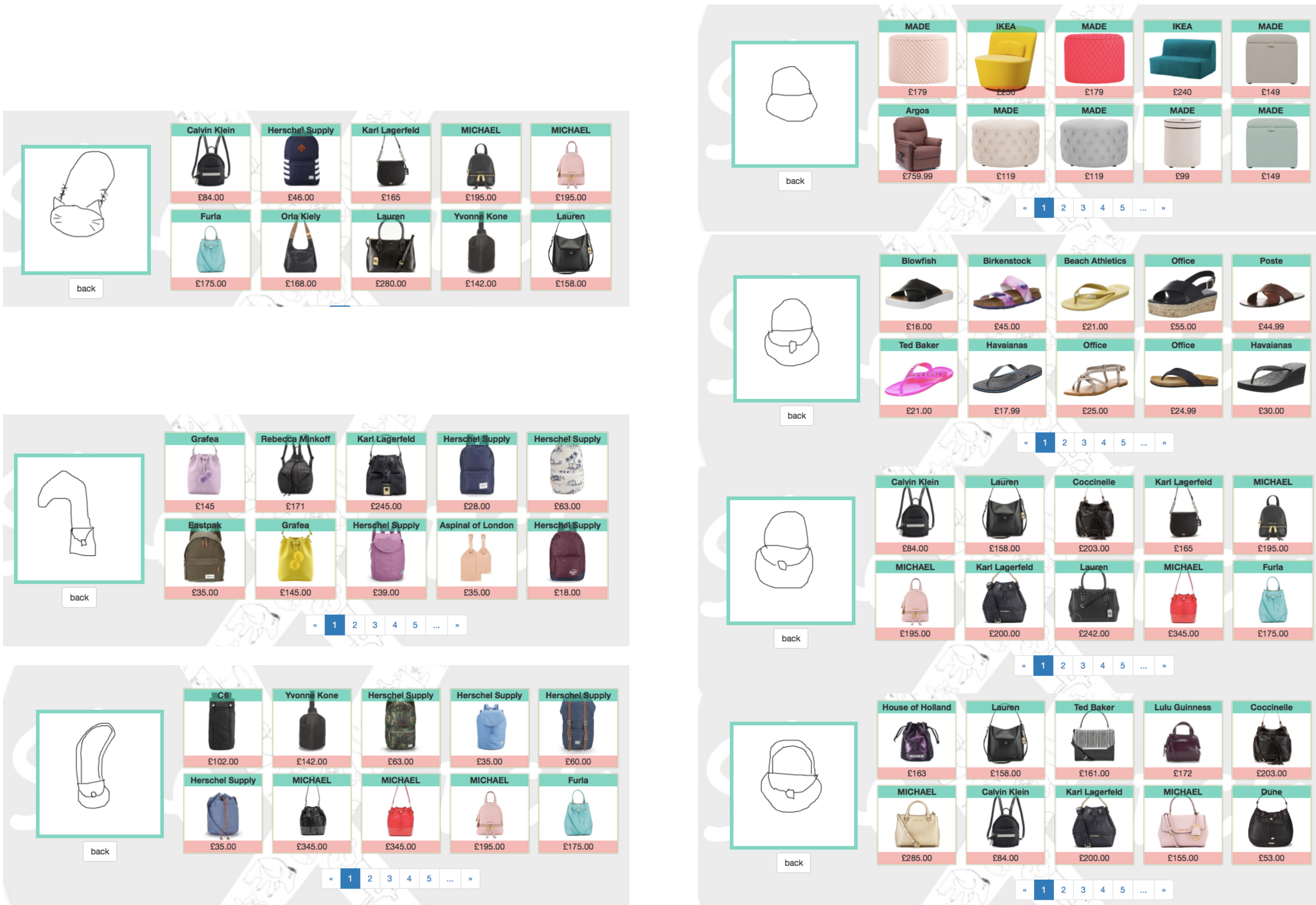
Conclusion

- 1st work to do fine-grained instance-level SBIR
- Limited amount of training data
    - Siamese network, triplet ranking
    - with more photo/sketch pair data, heterogeneous could be better

Demo


Kickers £75.00
Kickers £74.99
Vans California £25.00
Poste £68.00
Kickers £50.00
Poste £42.00
Vans £22.00
Vans £54.99
Vans £22.00
Timberland £55.00
back
« 1 2 3 4 5 ... »
Office £48.00
Office £65.00
Office £65.00
Office £68.00
Office £25.00
Office £65.00
United Nude £70.00
Office £50.00
Office £65.00
Ted Baker £95.00
back
« 1 2 3 4 5 ... »
Office £55.00
Poste Mistress £98.00
Office £53.00
Office £81.00
Office £58.00
Office £38.00
Office £63.00
Office £51.00
Office £41.00
Office £70.00
back
« 1 2 3 4 5 ... »
MICHAEL £195.00
MICHAEL £195.00
Calvin Klein £84.00
Furla £175.00
Karl Lagerfeld £165
MICHAEL £345.00
Yvonne Kone £142.00
Kate Sheridan £90.00
MICHAEL £345.00
Furla £175.00
back
« 1 2 3 4 5 ... »
Office £31.00
Office £37.00
Bedroom Athletics £13.00
Timberland £88.00
UGG Australia £72.00
Shoe the Bear £51.00
Blowfish £34.00
Ted Baker £81.00
Office £19.00
Cortica £69.99
back
« 1 2 3 4 5 ... »

Demo



back
Calvin Klein
£84.00
Herschel Supply
£46.00
Karl Lagerfeld
£165
MICHAEL
£195.00
MICHAEL
£195.00
Furla
£175.00
Orla Kiely
£168.00
Lauren
£280.00
Yvonne Kone
£142.00
Lauren
£158.00
back
Grafea
£145
Rebecca Minkoff
£171
Karl Lagerfeld
£245.00
Herschel Supply
£28.00
Herschel Supply
£63.00
Eastpak
£35.00
Grafea
£145.00
Herschel Supply
£39.00
Aspinal of London
£35.00
Herschel Supply
£18.00
back
C6
£102.00
Yvonne Kone
£142.00
Herschel Supply
£63.00
Herschel Supply
£35.00
Herschel Supply
£60.00
Herschel Supply
£35.00
MICHAEL
£345.00
MICHAEL
£345.00
MICHAEL
£195.00
Furla
£175.00
back
MADE
£179
IKEA
MADE
£179
IKEA
£240
MADE
£149
Argos
£759.99
MADE
£119
MADE
£119
MADE
£99
MADE
£149
back
Blowfish
£16.00
Birkenstock
£45.00
Beach Athletics
£21.00
Office
£55.00
Poste
£44.99
Ted Baker
£21.00
Havaianas
£17.99
Office
£25.00
Office
£24.99
Havaianas
£30.00
back
Calvin Klein
£84.00
Lauren
£158.00
Coccinelle
£203.00
Karl Lagerfeld
£165
MICHAEL
£195.00
MICHAEL
£195.00
Karl Lagerfeld
£200.00
Lauren
£242.00
MICHAEL
£345.00
Furla
£175.00
back
House of Holland
£163
Lauren
£158.00
Ted Baker
£161.00
Lulu Guinness
£172
Coccinelle
£203.00
MICHAEL
£285.00
Calvin Klein
£84.00
Karl Lagerfeld
£200.00
MICHAEL
£155.00
Dune
£53.00

Demo

https://www.eecs.qmul.ac.uk/~qian/Project_cvpr16.html