

Sketch Me That Shoe

Qian Yu et al.

CVPR 2016

presenter: Wei-Lin Hsiao
advisor: Kristen Grauman



back





Traditional text-based image retrieval:



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

"A ladies shoe..."



Image retrieval by **text** is challenging



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

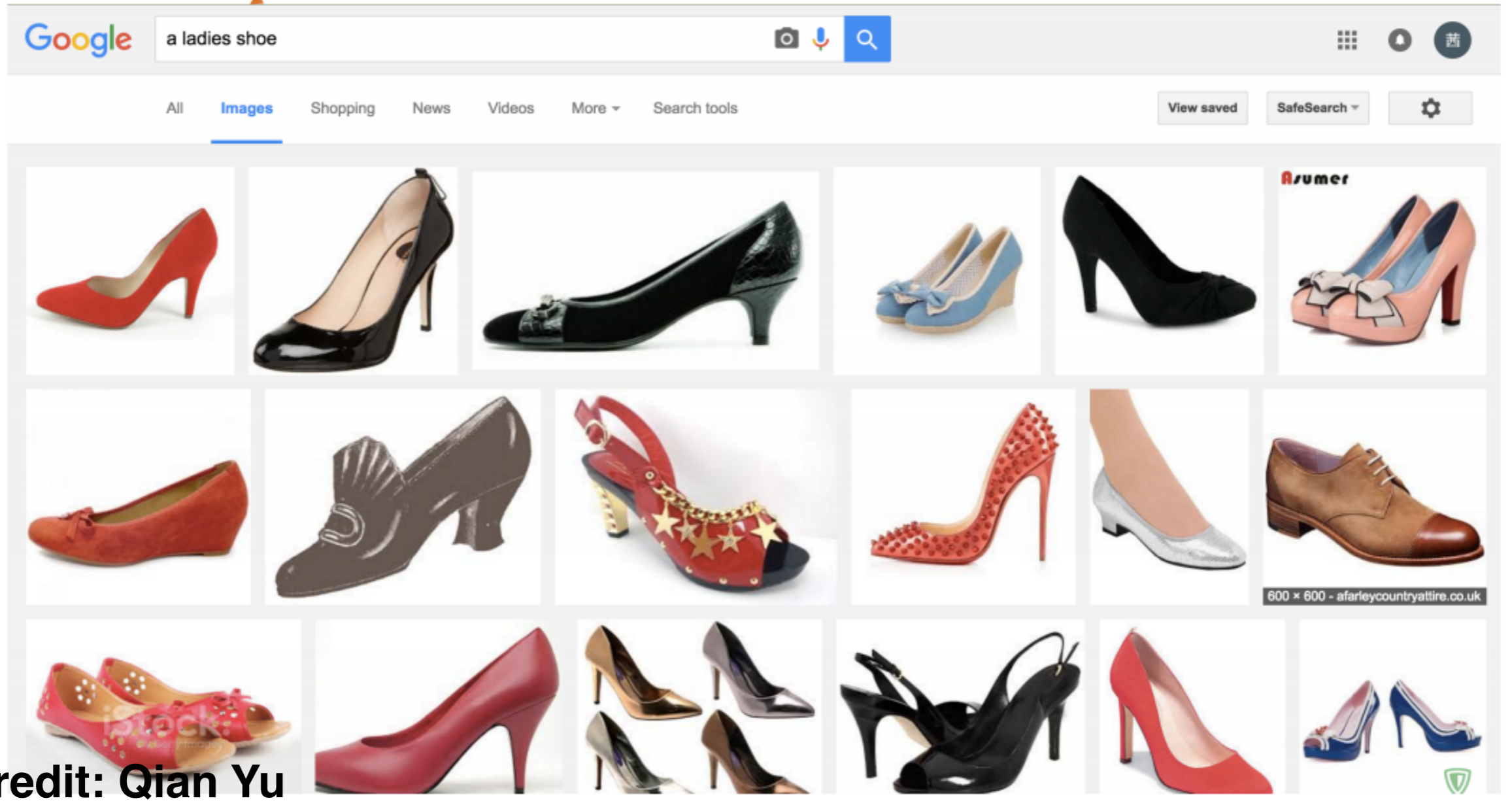
A screenshot of a Google Images search results page. The search bar contains the text "A sandal with open-toe, medium heel and lace". Below the search bar, there are navigation tabs for "All", "Shopping", "Images", "Videos", "News", and "More". The "Images" tab is selected. The search results are displayed as a grid of 24 images, showing various styles of sandals, including high-heeled, wedge, and lace-up designs. The images are arranged in three rows of eight. The first row shows a variety of styles, including black high-heeled sandals, white lace-up sandals, and blue and pink sneakers. The second row shows black lace-up sandals, white lace-up sandals, black lace-up sandals, white lace-up sandals, black lace-up sandals, and white lace-up sandals. The third row shows tan lace-up sandals, black lace-up sandals, black lace-up sandals, black lace-up sandals, black lace-up sandals, white lace-up sandals, and tan lace-up sandals.

slide credit: Qian Yu

Image retrieval by **text** is challenging



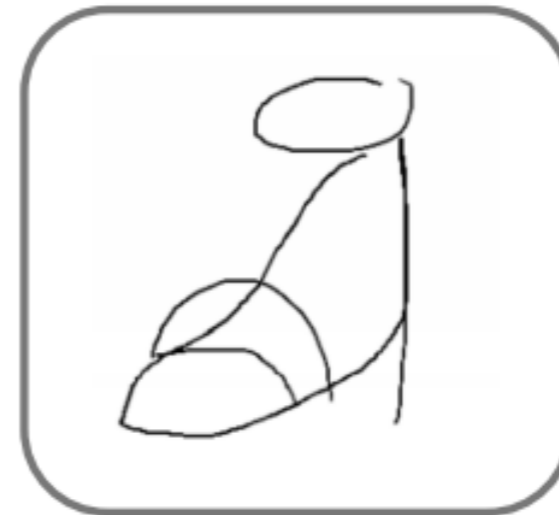
"A ladies shoe"



The screenshot shows a Google search interface. The search bar contains the text "a ladies shoe". Below the search bar, there are tabs for "All", "Images", "Shopping", "News", "Videos", and "More". The "Images" tab is selected. The search results are displayed in a grid of 18 images, showing various styles of ladies' shoes, including pumps, loafers, and sandals. The images are arranged in three rows of six. The first row shows a red pump, a black pump, a black pump with a bow, a blue loafer, a black pump, and a pink loafer. The second row shows a red loafer, a black loafer, a red loafer with gold stars, a red loafer with studs, a silver loafer, and a brown loafer. The third row shows a pink loafer, a red pump, a black pump, a black pump, a red pump, and a blue loafer. The search bar also includes icons for camera, voice search, and search. The Google logo is visible on the left. The search bar also contains the text "a ladies shoe".

slide credit: Qian Yu








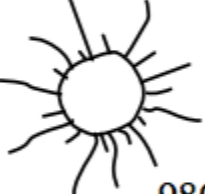










A sketch speaks for a hundred words



Sketch-based image retrieval (SBIR) — related work

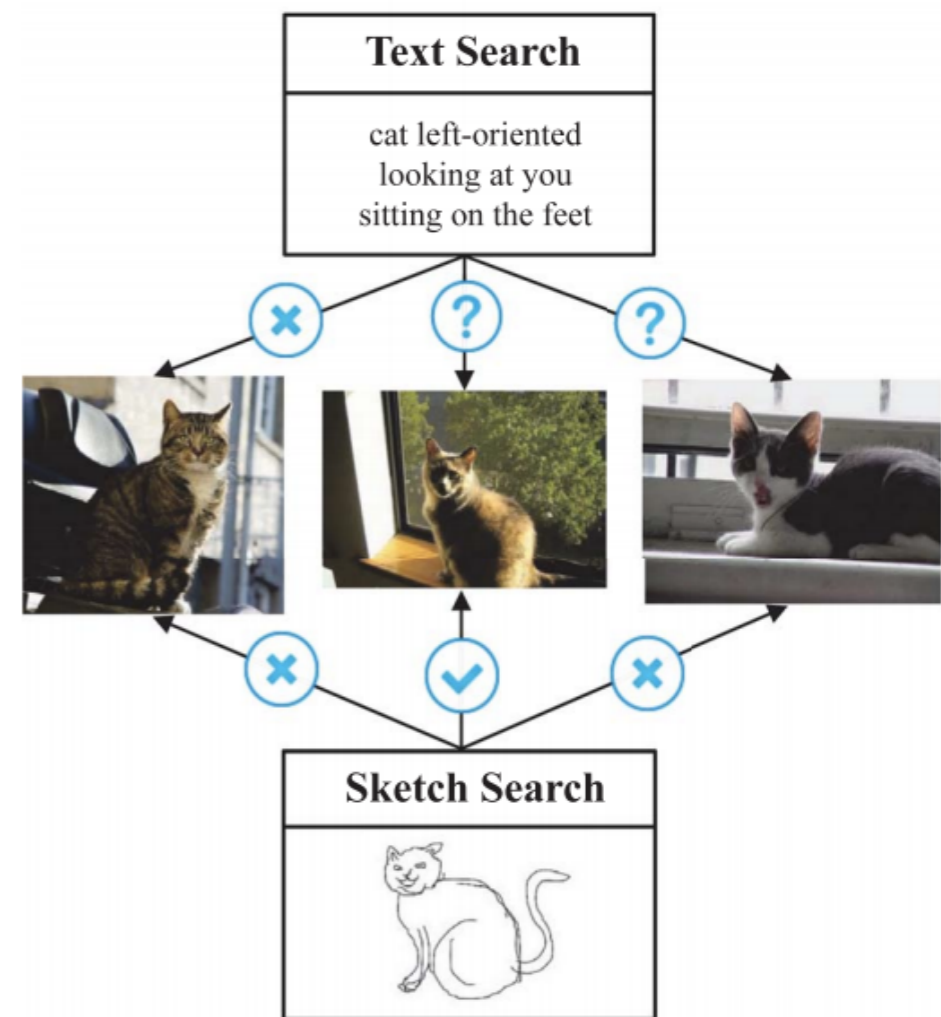
- **Category-level SBIR:**

- E. Mathis et al. TVCG 2011, E. Mathis et al. Computers & Graphics 2010, R. Hu ICIP 2010, Y. Cao, ACM 2010,

t-shirt  100%	snake  99%	comb  99%	flower  99%	eyeglasses  98%	elephant  98%
leaf  98%	sun  98%	wrist-watch  96%	pineapple  96%	trousers  96%	ladder  96%
apple  96%	airplane  96%	butterfly  96%	umbrella  96%	chair  95%	key  95%

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

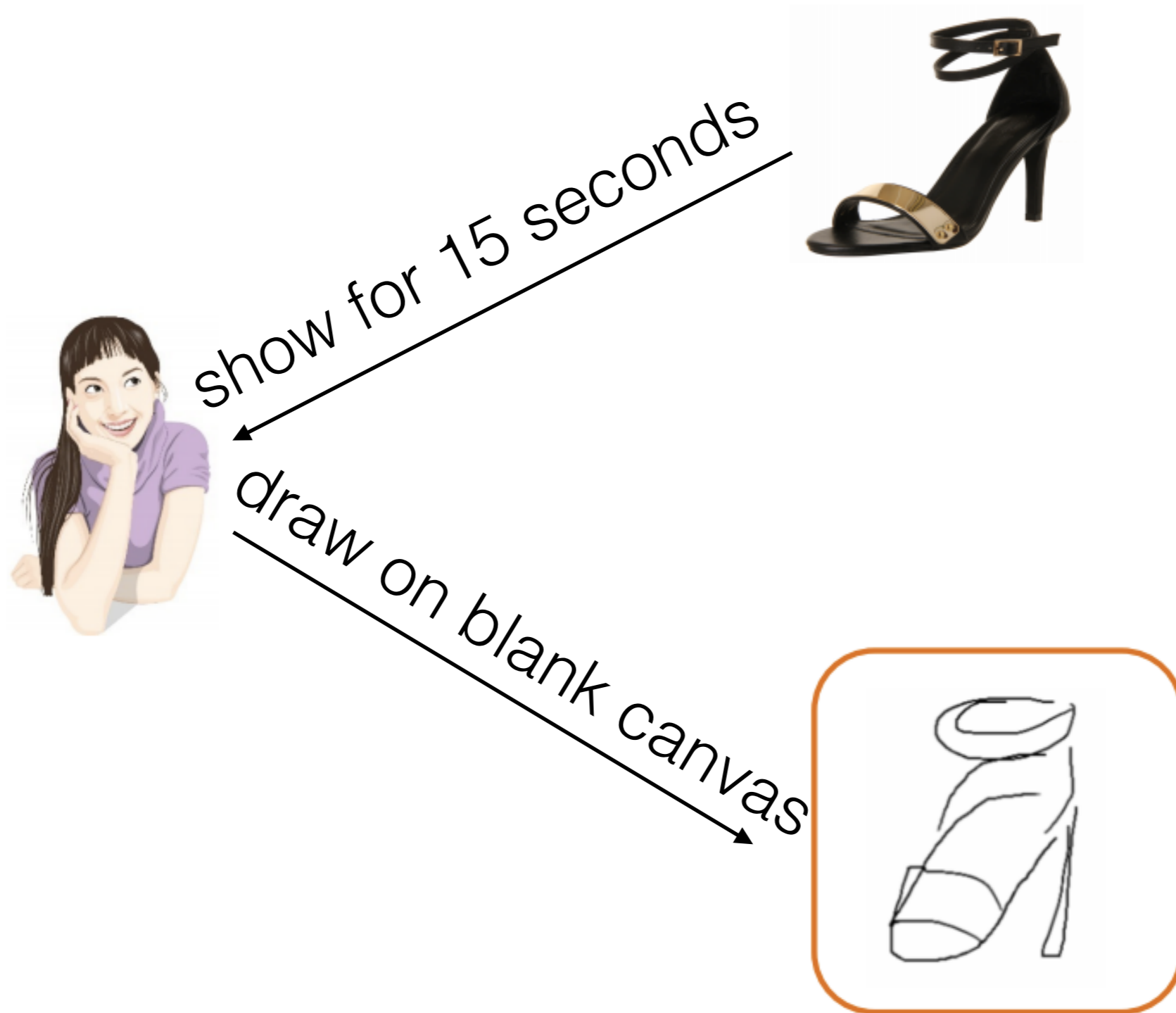
2. Overcome the requirements of extensive data and annotation by

- 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:

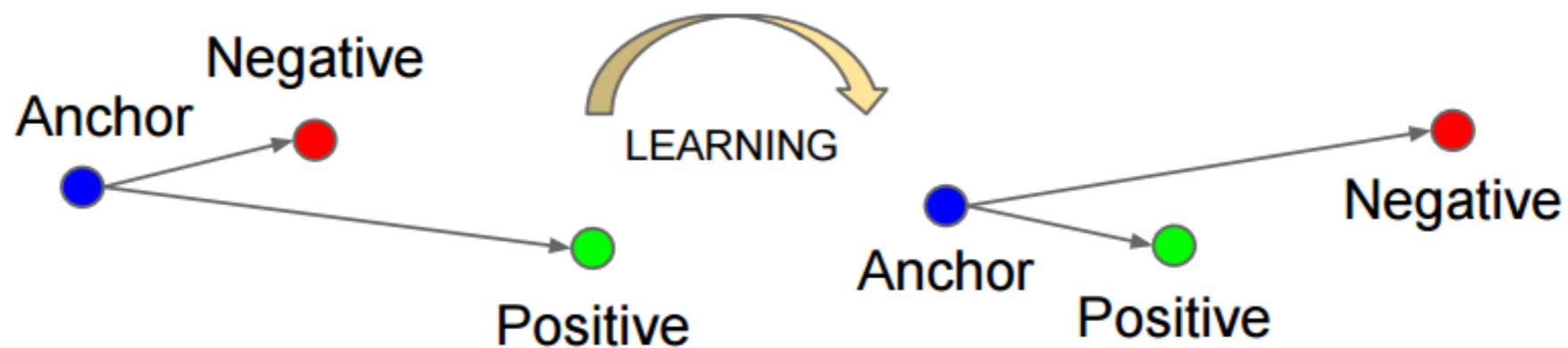
- Top 10 closest photo images to the query sketch

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

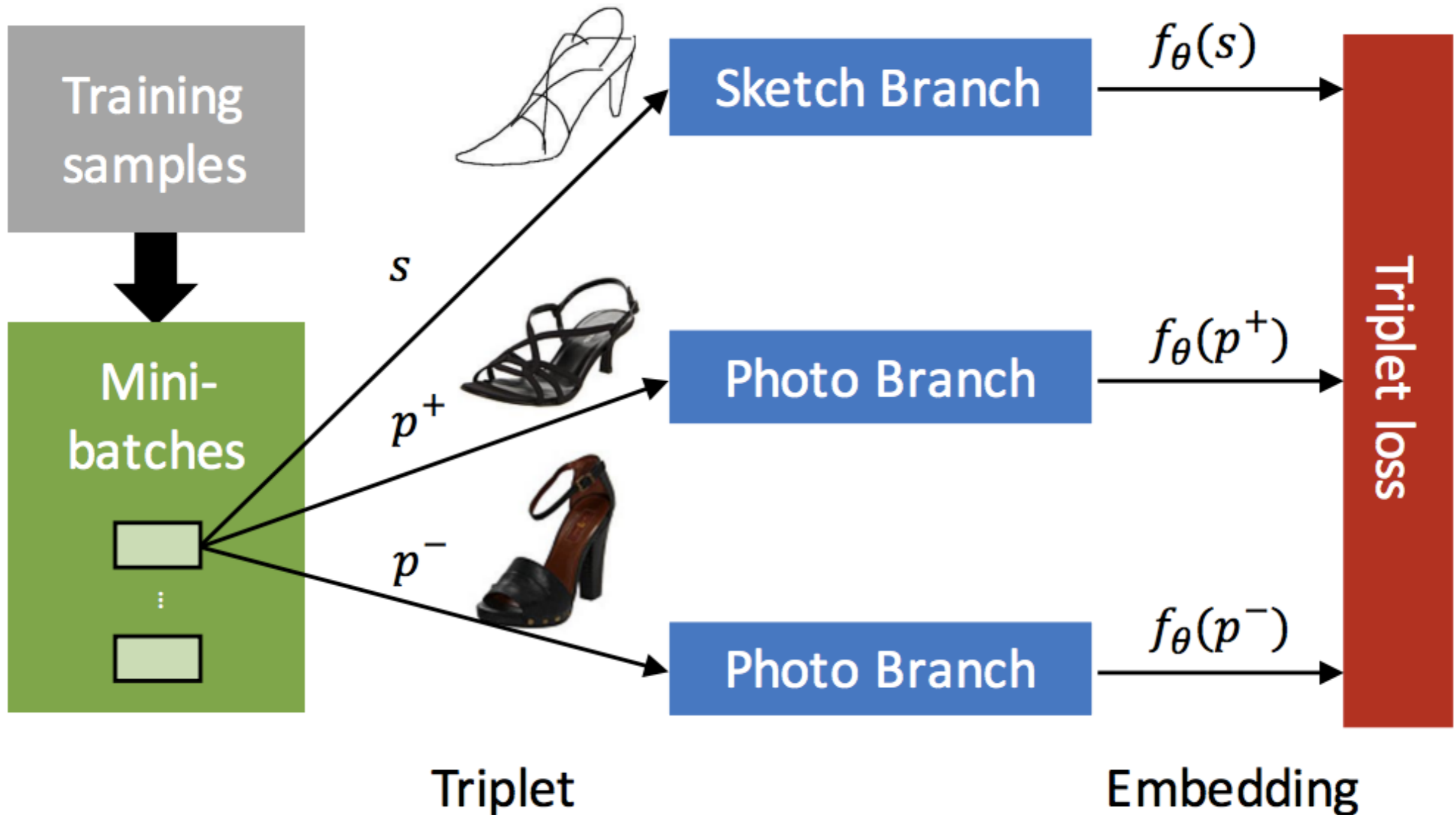
$$\min_{\theta} \sum_{t \in T} L_{\theta}(t) + \lambda R(\theta)$$



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

distance between sketch and positive photo - distance between sketch and negative photo

Network architecture



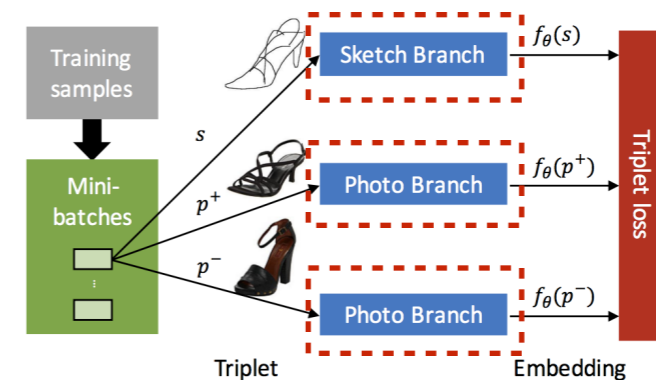
Pre-train/fine-tune

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— Step 1,2

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

1. Photos:



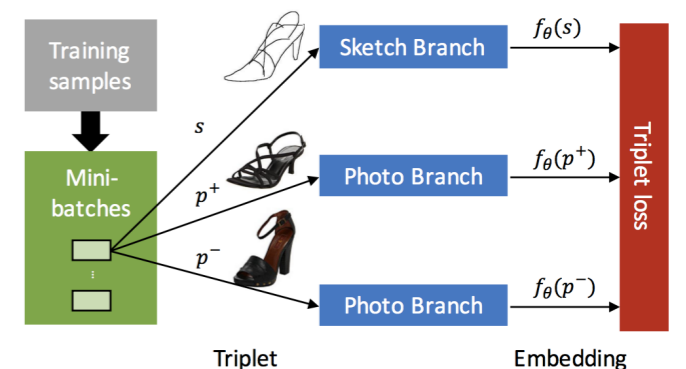
- Pre-train to classify 1000 categories of ImageNet-1K with **edge maps** extracted

2. Free-hand sketches:

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

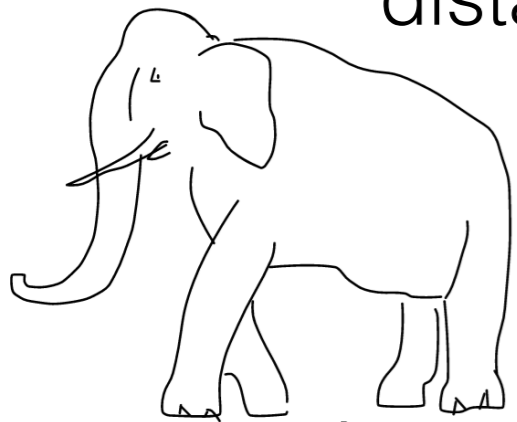
Exploit auxiliary sketch/photo category-paired data—Step 3

- Train sketch-photo ranking network:
 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



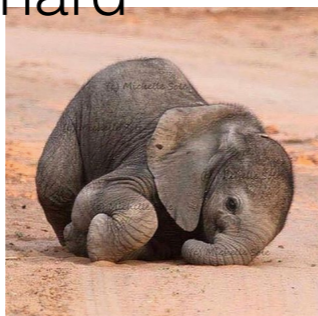
random
different classes

out-of-class hard



distances smaller than positives
different classes

in-class hard



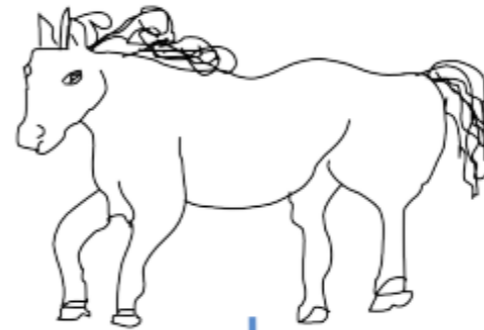
bottom 20% most similar
same class

Fine-tune on target scenario —Step 4

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

Data augmentation

Input
(1 Sketch)

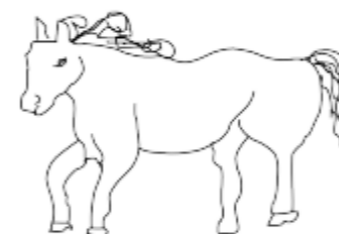


Stroke Removal
(3 sketches)

remove 10%

remove 30%

remove 50%



shorter and **later** strokes more likely to be removed

Stroke Deformation
(9 sketches)



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

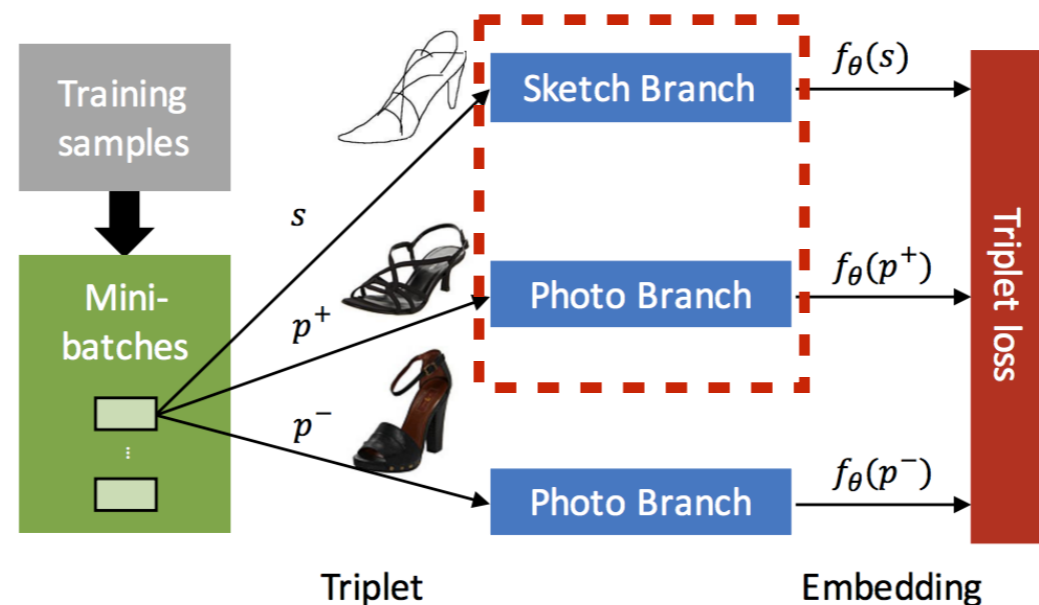


Contribution of different component

	acc.@1	acc.@10
without any pretraining		
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?

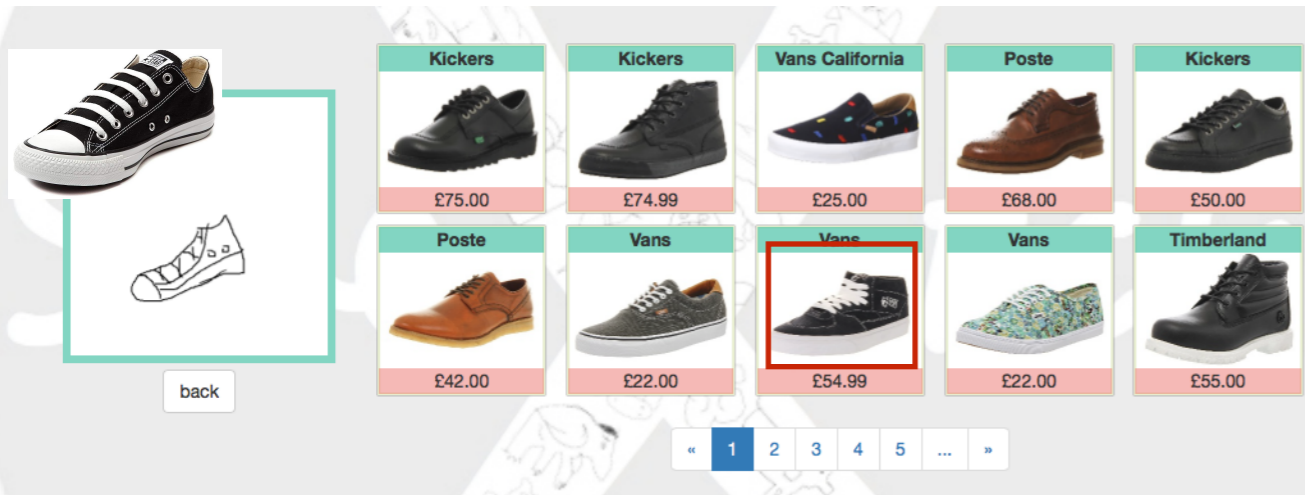
	acc.@1	acc.@10
Siamese verification	28.70%	78.26%
Hetero. ranking	21.74%	68.70%
Hetero. verification	16.52%	69.57%
Our full model <small>siamese, ranking</small>	39.13%	87.83%



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

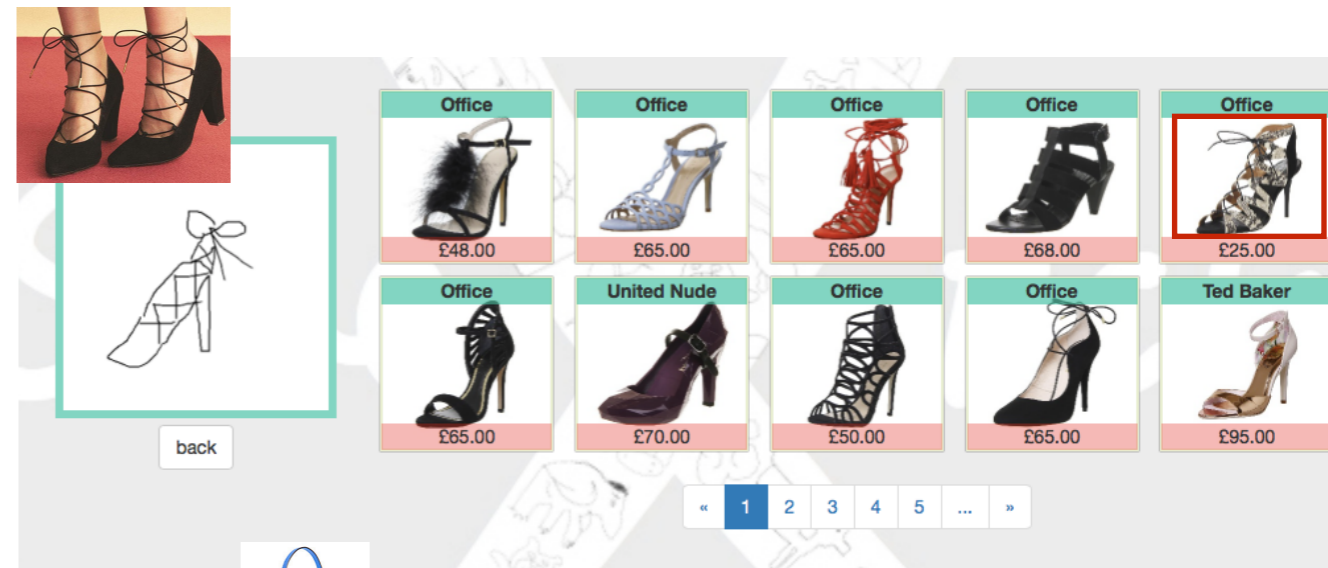
Demo



Product grid for sneakers. The selected item is a black and white sneaker. The grid shows various styles from brands like Kickers, Vans, Poste, and Timberland. A 'back' button is visible below the selected item.

£75.00	£74.99	£25.00	£68.00	£50.00	
£42.00	£22.00	£54.99	£22.00	£55.00	

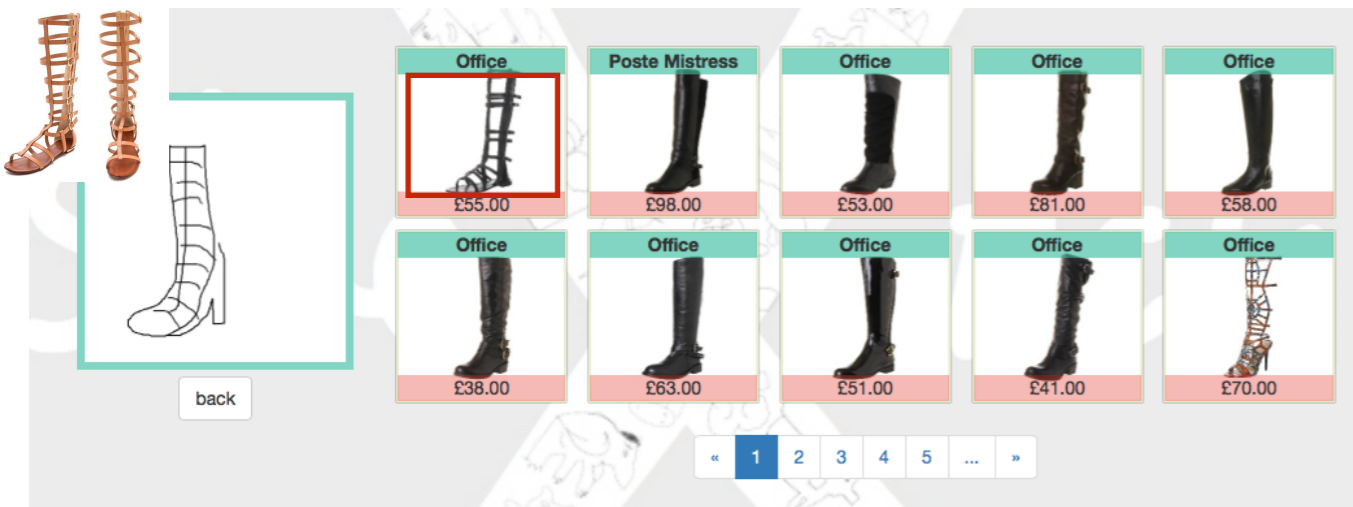
« 1 2 3 4 5 ... »



Product grid for high-heeled shoes. The selected item is a black lace-up high-heeled shoe. The grid shows various styles from brands like Office, United Nude, and Ted Baker. A 'back' button is visible below the selected item.

£48.00	£65.00	£65.00	£68.00	£25.00	
£65.00	£70.00	£50.00	£65.00	£95.00	

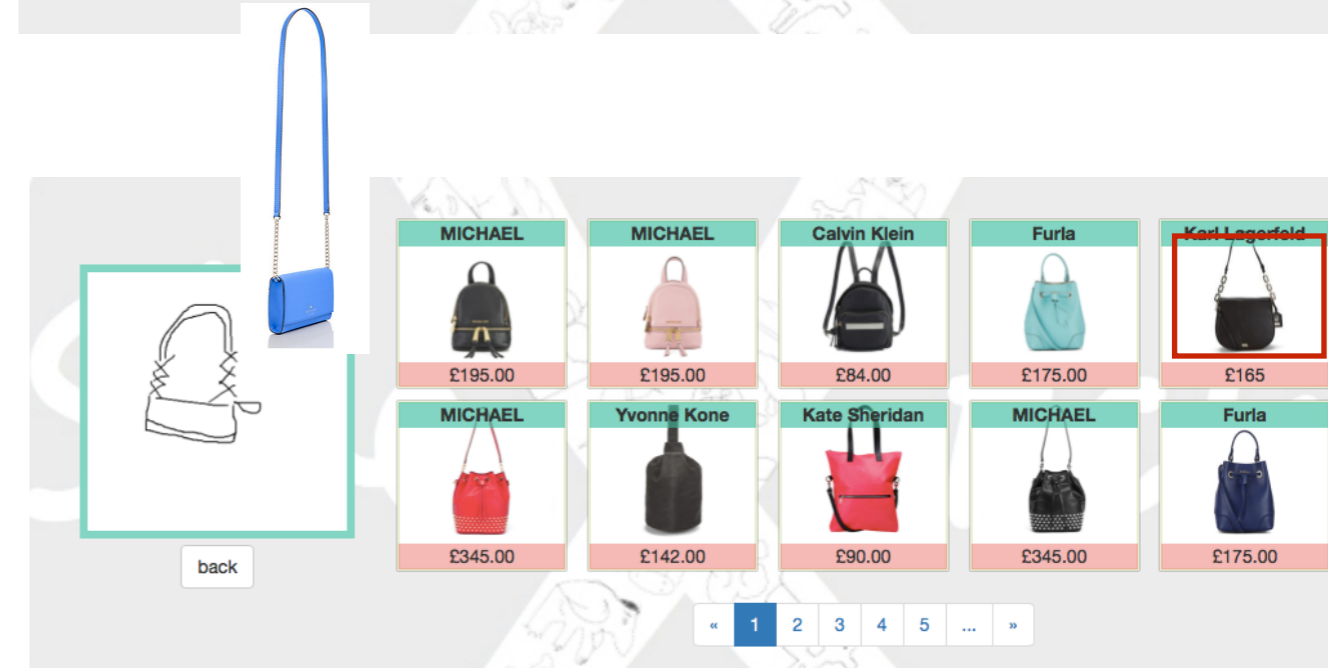
« 1 2 3 4 5 ... »



Product grid for boots. The selected item is a black lace-up boot. The grid shows various styles from brands like Office and Poste. A 'back' button is visible below the selected item.

£55.00	£98.00	£53.00	£81.00	£58.00	
£38.00	£63.00	£51.00	£41.00	£70.00	

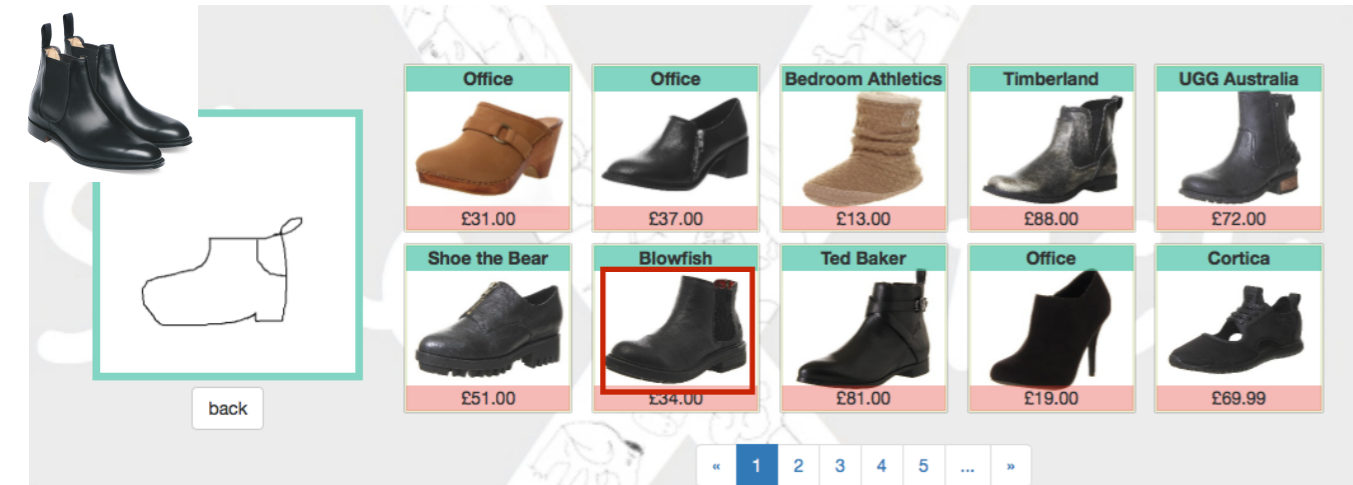
« 1 2 3 4 5 ... »



Product grid for handbags. The selected item is a black handbag. The grid shows various styles from brands like Michael Kors, Calvin Klein, Furla, and Karl Lagerfeld. A 'back' button is visible below the selected item.

£195.00	£195.00	£84.00	£175.00	£165	
£345.00	£142.00	£90.00	£345.00	£175.00	

« 1 2 3 4 5 ... »



Product grid for boots. The selected item is a black Chelsea boot. The grid shows various styles from brands like Office, Bedroom Athletics, Timberland, UGG Australia, Shoe the Bear, Blowfish, Ted Baker, Office, and Cortica. A 'back' button is visible below the selected item.

£31.00	£37.00	£13.00	£88.00	£72.00	
£51.00	£34.00	£81.00	£19.00	£69.99	

« 1 2 3 4 5 ... »

Demo



back

Calvin Klein	Herschel Supply	Karl Lagerfeld	MICHAEL	MICHAEL
£84.00	£46.00	£165	£195.00	£195.00
Furla	Orla Kiely	Lauren	Yvonne Kone	Lauren
£175.00	£168.00	£280.00	£142.00	£158.00

back

Grafea	Rebecca Minkoff	Karl Lagerfeld	Herschel Supply	Herschel Supply
£145	£171	£245.00	£28.00	£63.00
Eastpak	Grafea	Herschel Supply	Aspinal of London	Herschel Supply
£35.00	£145.00	£39.00	£35.00	£18.00

back

CB	Yvonne Kone	Herschel Supply	Herschel Supply	Herschel Supply
£102.00	£142.00	£63.00	£35.00	£60.00
Herschel Supply	MICHAEL	MICHAEL	MICHAEL	Furla
£35.00	£345.00	£345.00	£195.00	£175.00

back

MADE	IKEA	MADE	IKEA	MADE
£179	£250	£179	£240	£149
Argos	MADE	MADE	MADE	MADE
£759.99	£119	£119	£99	£149

back

Blowfish	Birkenstock	Beach Athletics	Office	Poste
£16.00	£45.00	£21.00	£55.00	£44.99
Ted Baker	Havaianas	Office	Office	Havaianas
£21.00	£17.99	£25.00	£24.99	£30.00

back

Calvin Klein	Lauren	Coccinelle	Karl Lagerfeld	MICHAEL
£84.00	£158.00	£203.00	£165	£195.00
MICHAEL	Karl Lagerfeld	Lauren	MICHAEL	Furla
£195.00	£200.00	£242.00	£345.00	£175.00

back

House of Holland	Lauren	Ted Baker	Lulu Guinness	Coccinelle
£163	£158.00	£161.00	£172	£203.00
MICHAEL	Calvin Klein	Karl Lagerfeld	MICHAEL	Dune
£285.00	£84.00	£200.00	£155.00	£53.00

Demo

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