Sketch Me That Shoe

Heechan Shin

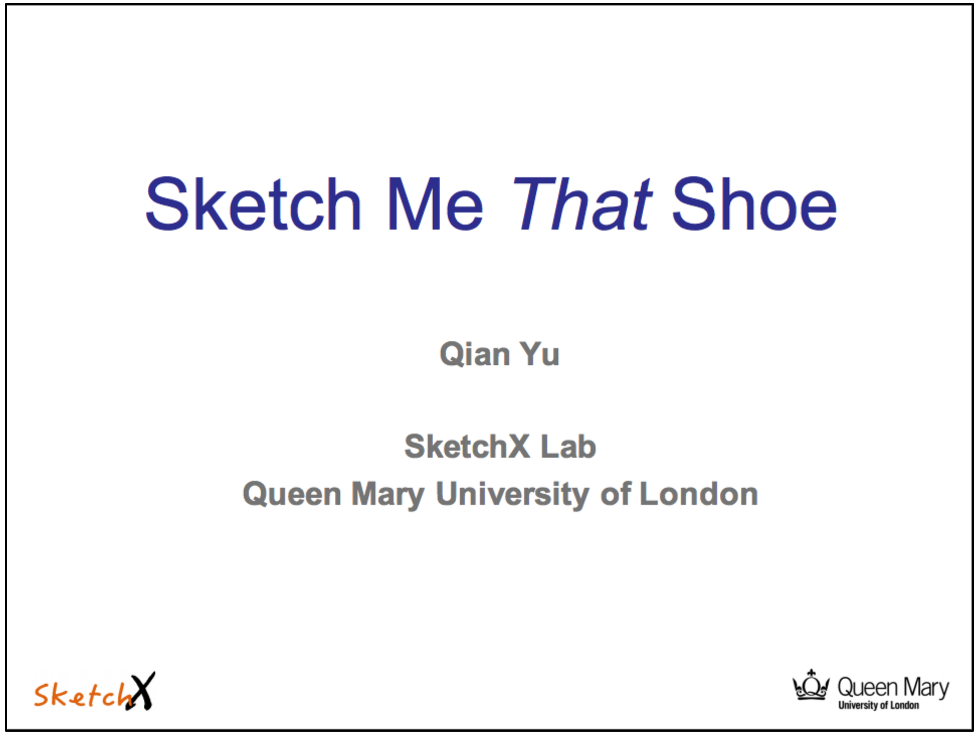
CS688
Student paper presentation

Contents

- **Problems**
- **Solution**
  - Dataset
  - Methodology
- **Experiment**

Announcement

- Most of contents of this presentation comes from materials of author's CVPR presentation.



Problems

- Sketch Based Image Retrieval (SBIR)

Problems

- SBIR
  - Pros
    - No need for complicated description
    - No need for photos
  - Cons
    - Sketch is highly abstract
    - Heterogeneous domains (sketch ↔ image)

Problems

- Previous works
  - Eitz, Mathias, et al. *"An evaluation of descriptors for large-scale image retrieval from sketched feature lines."* Computers & Graphics, 2010
  - Eitz, Mathias, et al. *"Sketch-based image retrieval: Benchmark and bag-of-features descriptors."* TVCG, 2011
  - Hu, Rui, et al. "Gradient field descriptor for sketch based retrieval and localization." ICIP, 2010

⇒ Category-level SBIR

Problems

vs.

Category-level SBIR

Instance-level SBIR

This work wants to find **fine-grained instance-level** SBIR

Problems

- Sketch Based Image Retrieval (SBIR)
  - Sketch
    - Edge maps (automatically generated)

    - Professional drawings (skilled artist)

    - **Free-hand sketches** (amateur)

Problems

- Reasons of challenging
  - Sketch is highly abstract
  - Heterogeneous domains (sketch ↔ image)

  } Cons of SBIR

  - Want to capture the **fine-grained** similarities with free-hand sketches
  - No large-scale dataset exists

Solutions

- Contributions
  - Constructing fine-grained SBIR dataset

  - Pre-training with sketch-specific data augmentation

Solutions

- Constructing fine-grained SBIR dataset
  1. Data collection
     1) Collecting photo images
        - 419 shoe images from UT-Zap50K, 297 chairs from IKEA, Amazon and Taobao
     2) Collecting sketches
        - Recruiting 22 volunteers

Solutions

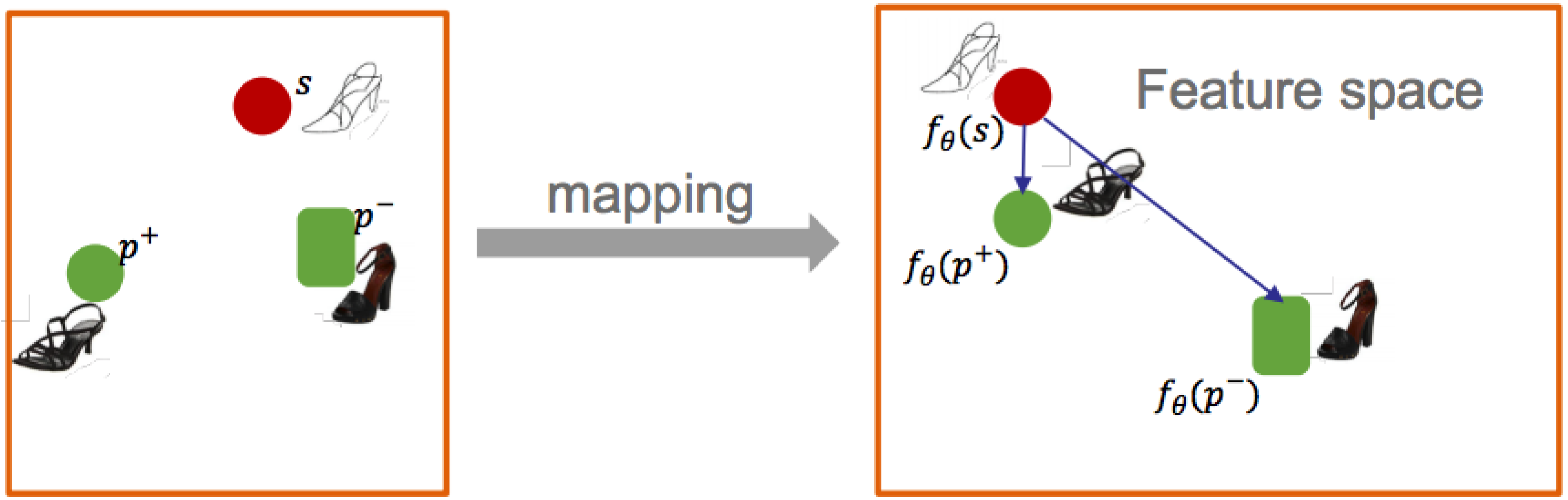
- Constructing fine-grained SBIR dataset
  2. Data annotation
     1) Attribute annotation
     2) Generating candidate photos for each sketch
     3) Triplet annotation

Solutions

- Learn a feature space using triplet loss
  - Always, $D\big(f_\theta(s), f_\theta(p^+)\big) < D\big(f_\theta(s), f_\theta(p^-)\big)$
  - Loss function :

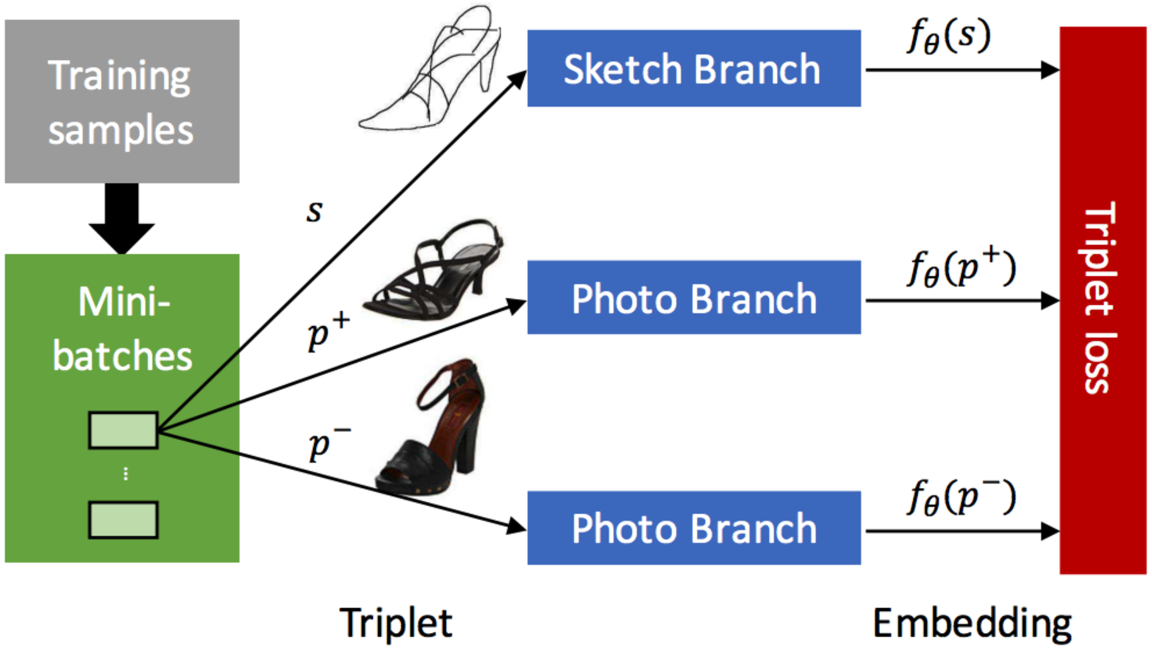
$$L_\theta(s, p^+, p^-) = \max\Big(0, \Delta + D\big(f_\theta(s), f_\theta(p^+)\big) - D\big(f_\theta(s), f_\theta(p^-)\big)\Big)$$

Where, $D(\cdot)$ is euclidean distance, $f_\theta(\cdot)$ is feature embedding function



Solutions

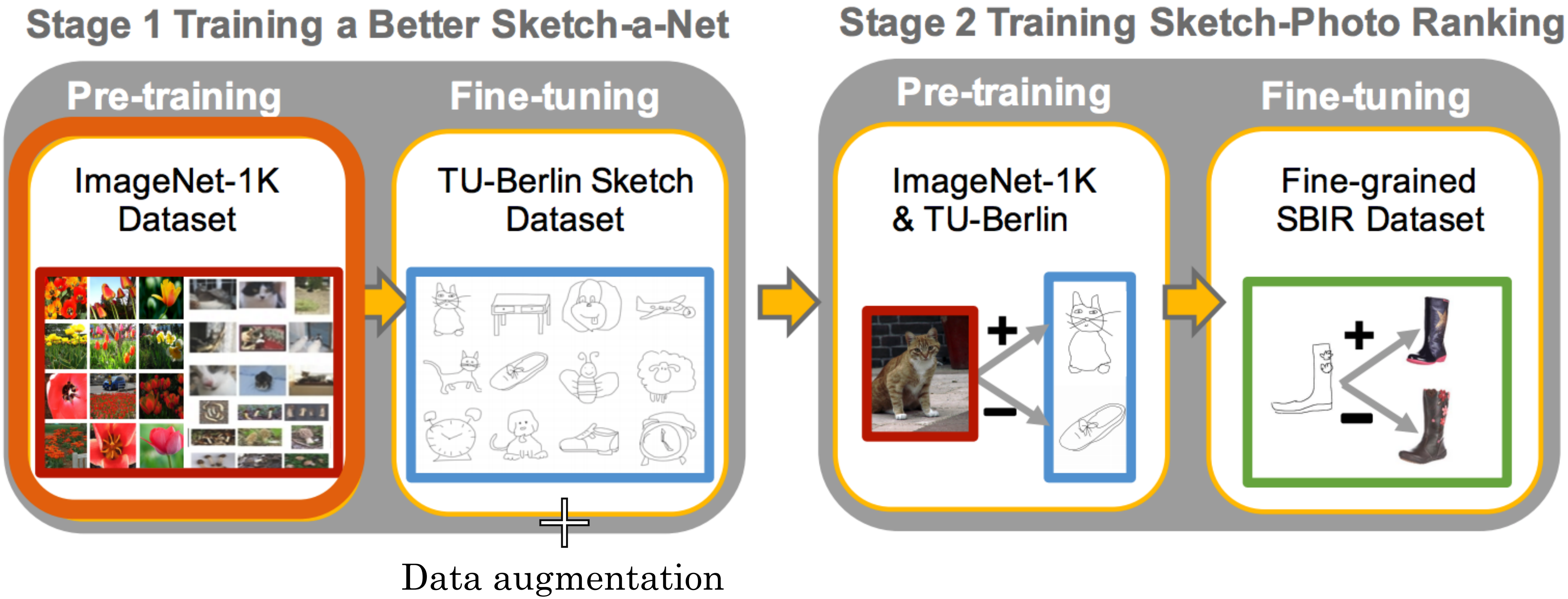
- Using three identical Sketch-a-Net* CNNs with Siamese network approach



* Q. Yu, et. al., *"Sketch-a-net that beats humans"* BMVC, 2015

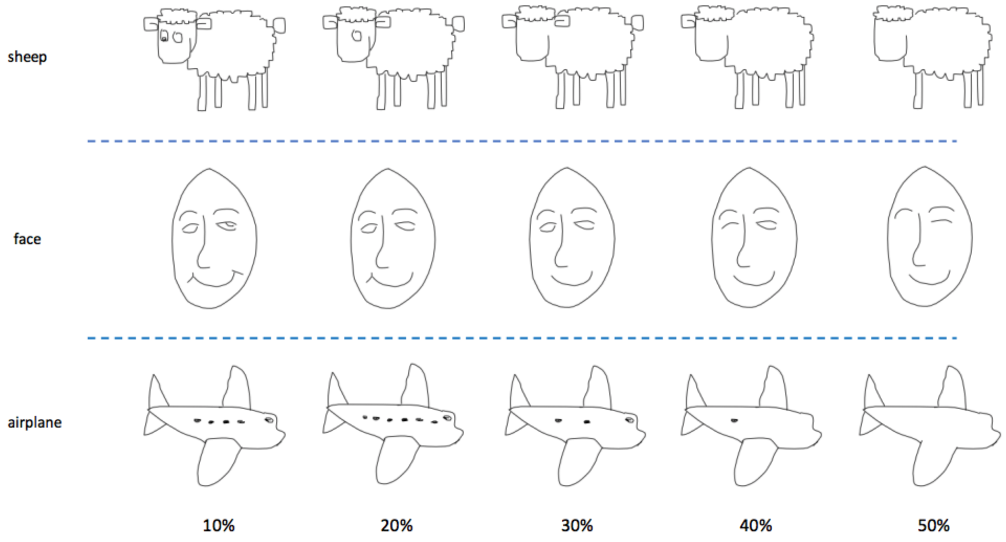
Solutions

- Re-train each Sketch-a-Net with

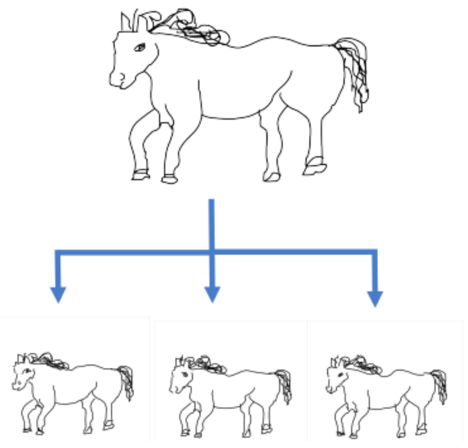


Solutions

- Data augmentation
  - Stroke removal
    - Broad outline is important
    - Longer line is important
    - Sketch is drawn from outside

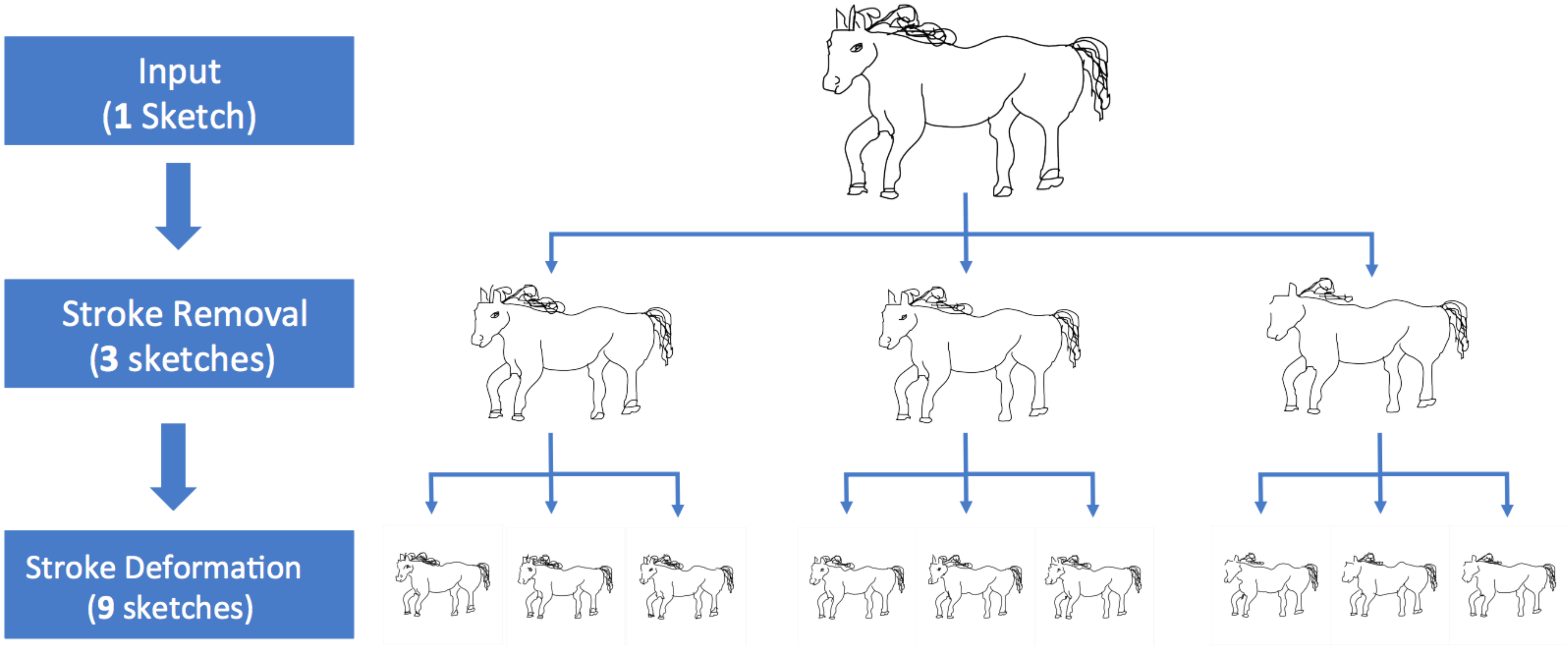


  - Stroke deformation
    - Using Moving Least Square algorithm



Solutions

- Data augmentation



Experiment

- Settings
  - Data
    - 419 shoes (304 for training + 115 for testing)
    - 297 chairs (200 for training + 97 for testing)
  - Implementation setting
    - Caffe
    - 32 CPU with 2 Nvidia Tesla K80
    - Learning rate : 0.001
    - Batch size : 128
    - During training, randomly crop 225×225 sub-images and flip them with 0.5 probability

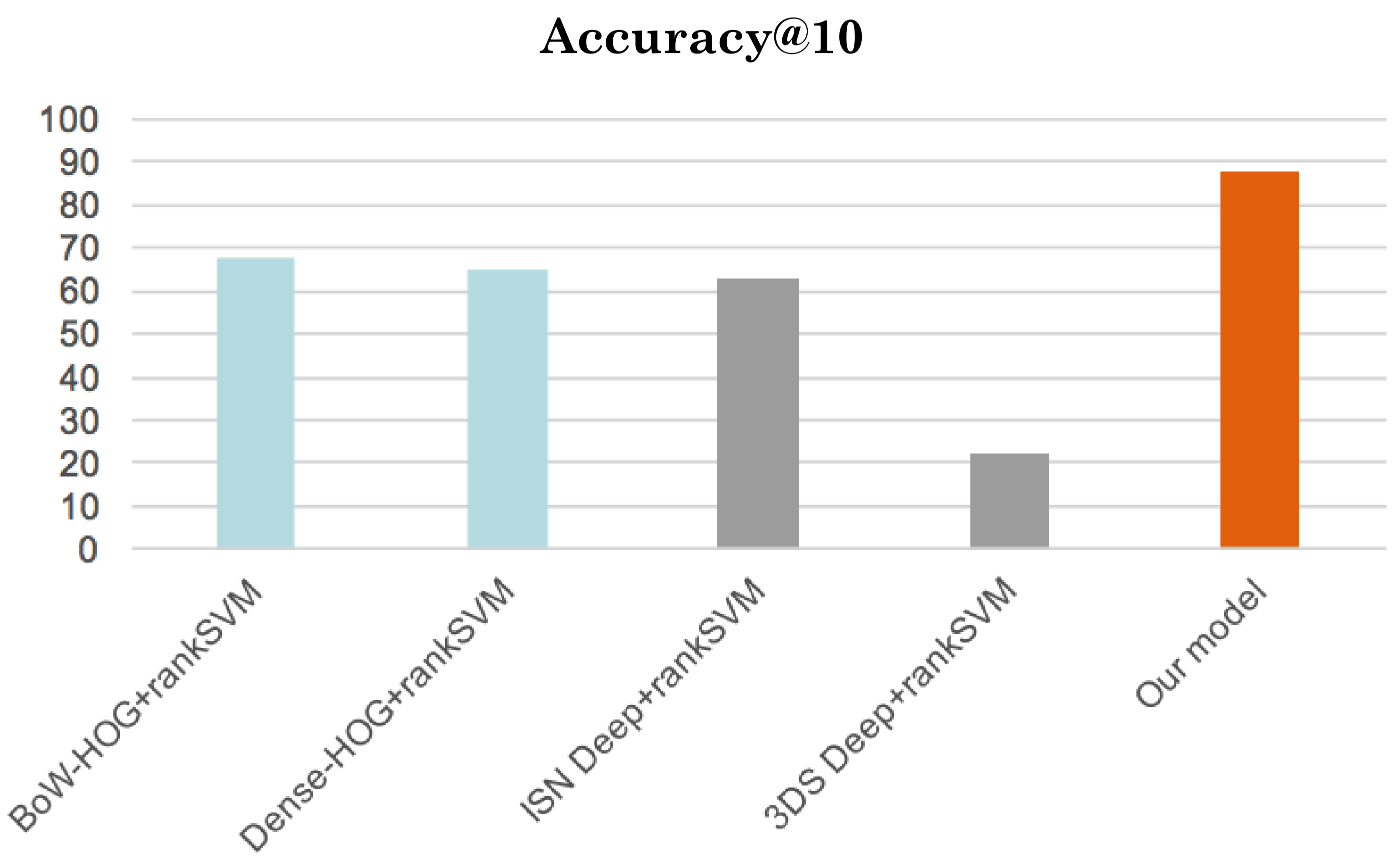
Experiment

Triplet-ranking prediction

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%**

Experiment


Accuracy@10
100
90
80
70
60
50
40
30
20
10
0
BoW-HOG+rankSVM
Dense-HOG+rankSVM
ISN Deep+rankSVM
3DS Deep+rankSVM
Our model

Experiment

30ms per one retrieval

https://sketchx.eecs.qmul.ac.uk

Thank you

- Quiz
  1. Which is the target of this work?
     ① Category – level SBIR
     ② Instance – level SBIR
     ③ Siamese – level SBIR

  2. In the data augmentation section, what did they do?
     ① Region removal & region deformation
     ② Stroke removal & stroke deformation
     ③ Context removal & context deformation