Image Retrieval with Mixed Initiative and Multimodal Feedback, BMVC '18

CS688 Paper Presentation

2018/11/22

20174315 Chiwan Song

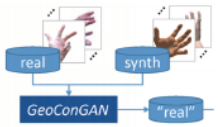
KAIST

GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB, Mueller et al., CVPR '18

Conclusion & Summary

- Presents a more robust model for occlusions
- Presents
  - a data set similar to the real hand domain
  - a model that can create the data set
- Demonstrates these benefits in the evaluation
  - particularly in difficult occlusion scenarios.
- Summary
  - Real-time full 3D hand tracking from single monocular RGB video.
  - Technical Novelties

1)

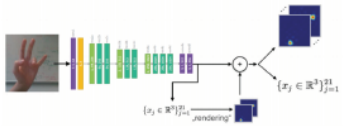


New GAN for **geometrically consistent unpaired image-to-image translation**

2)

Novel **enhanced RGB dataset** with **3D** hand joint **annotations** (>260k frames)

3)



CNN with **projection layer** for tightly coupled regression of **2D and 3D joint locations**

KAIST

Table of contents

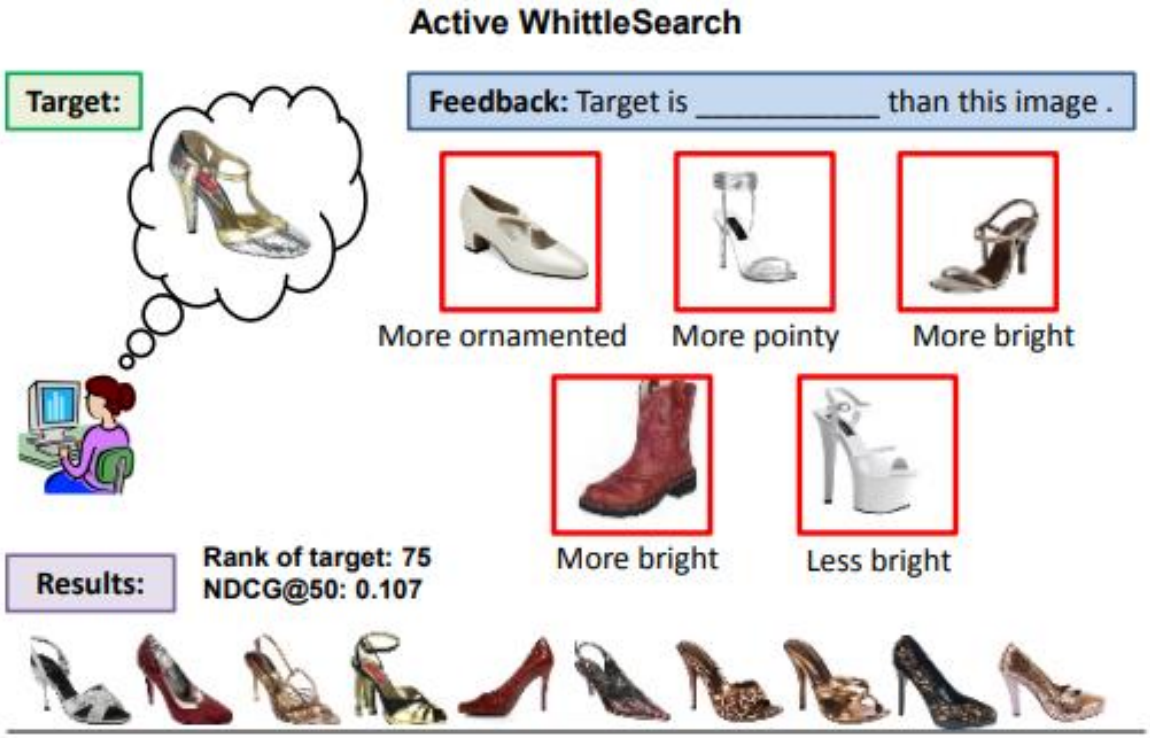
- Introduction
- Backgrounds
- Main idea
- Experiment & Result

KAIST

Introduction

Giving user's mental concept to a system

- The users can give language-based guidance to the system



5 Kovashka et al., IJCV '15

Giving user's mental concept to a system

- The system asks to users to derive what users want

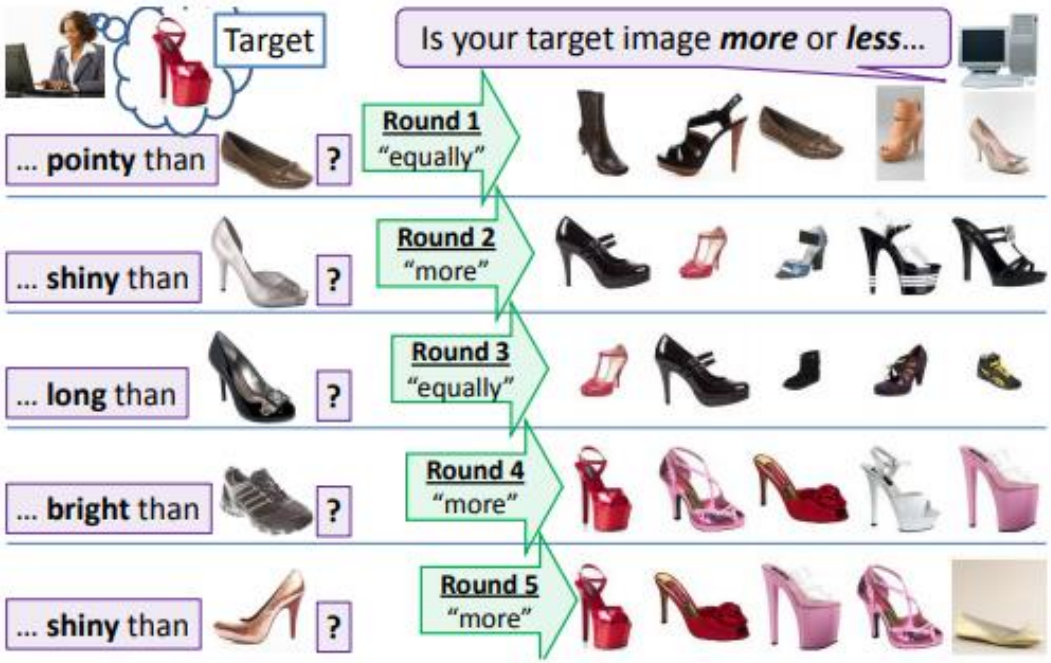


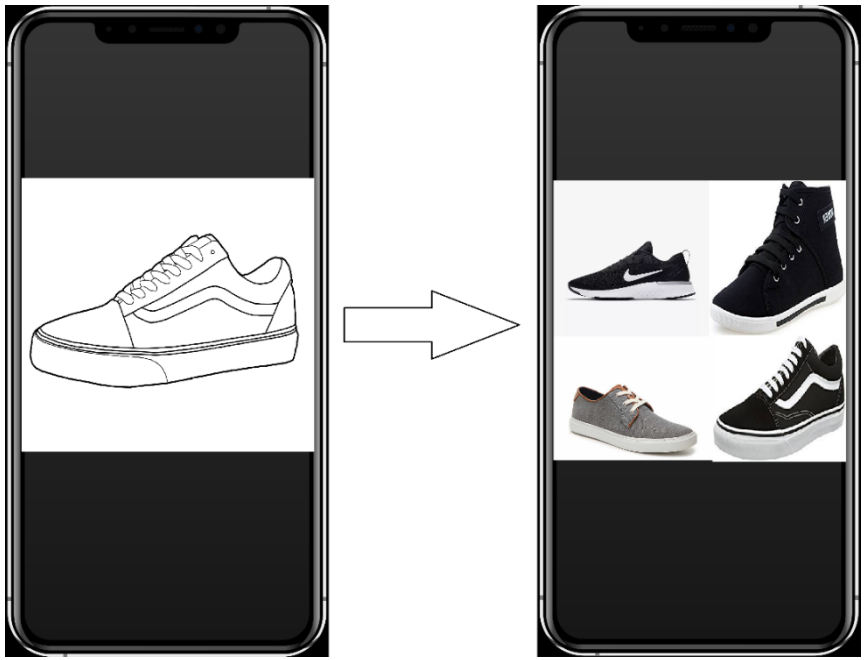
Figure 6. Using the user's feedback on the left, we retrieve the images on the right at the top of the results list.

 Kovashka et al., ICCV '13

KAIST

Giving user's mental concept to a system

- The users can provide visual cues for what they are looking for
  - i.e. Sketch based image retrieval



KAIST

Problems in previous work

- The interaction was driven by user **or** system, but **not both**
- The single modality of user input – language or visual

Main contribution of the paper

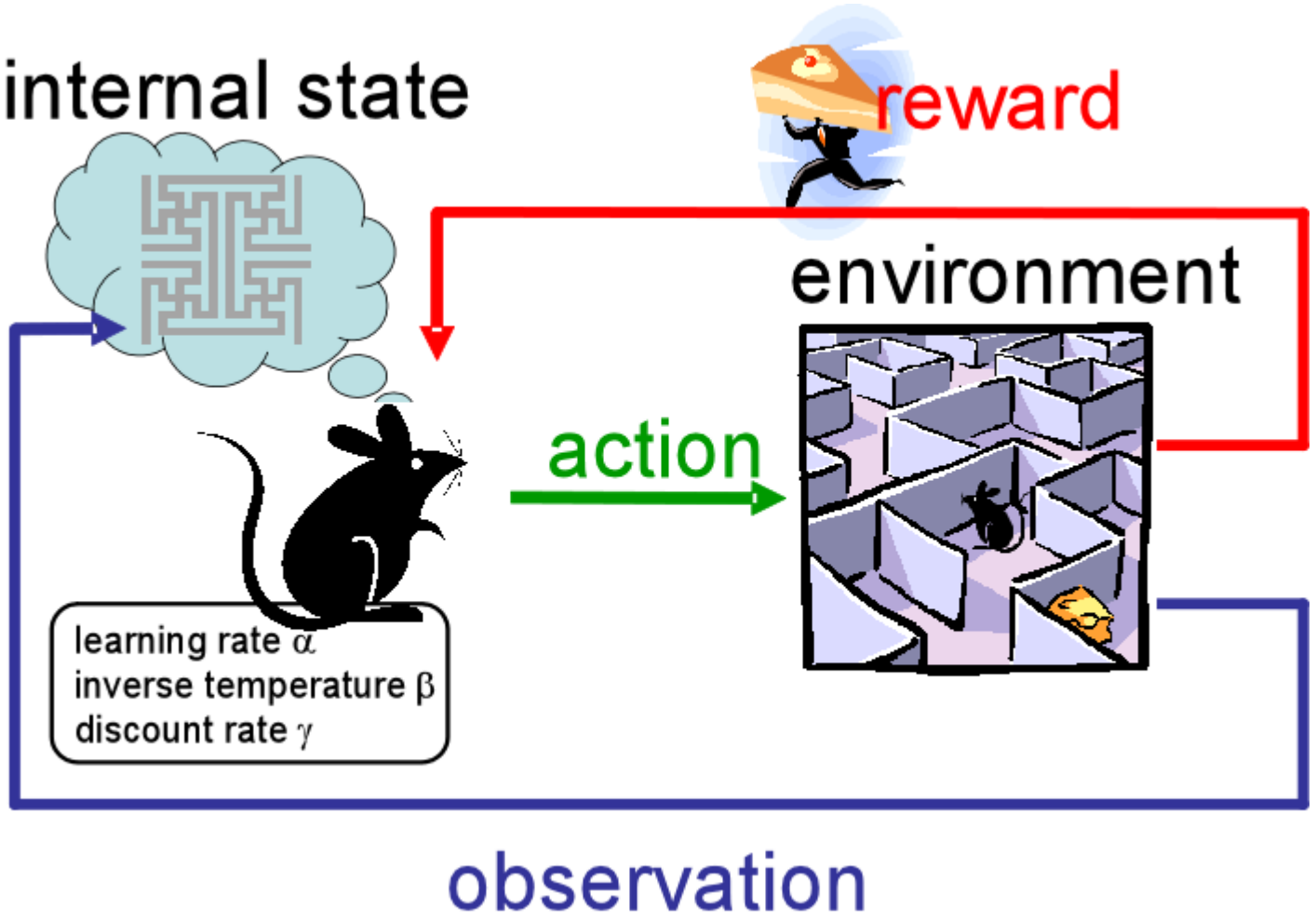
- A new framework that the interaction can be driven **either** the user **or** system
  - By agent trained with **Reinforcement Learning**
- The **dynamic decision** of interaction methods
  - Sketch, free-form attribute feedback, or answering attribute-based question

Backgrounds

Concept of Reinforcement Learning

- Inspired by the psychology of behavior
- Consist of **agent, environment, state, action, and reward**
- The **agent** changes **state** by doing **action**
- The **environment** gives **reward** to **agent** according to the **action**

Concept of Reinforcement Learning



[12] https://becominghuman.ai/the-very-basics-of-reinforcement-learning-154f28a79071

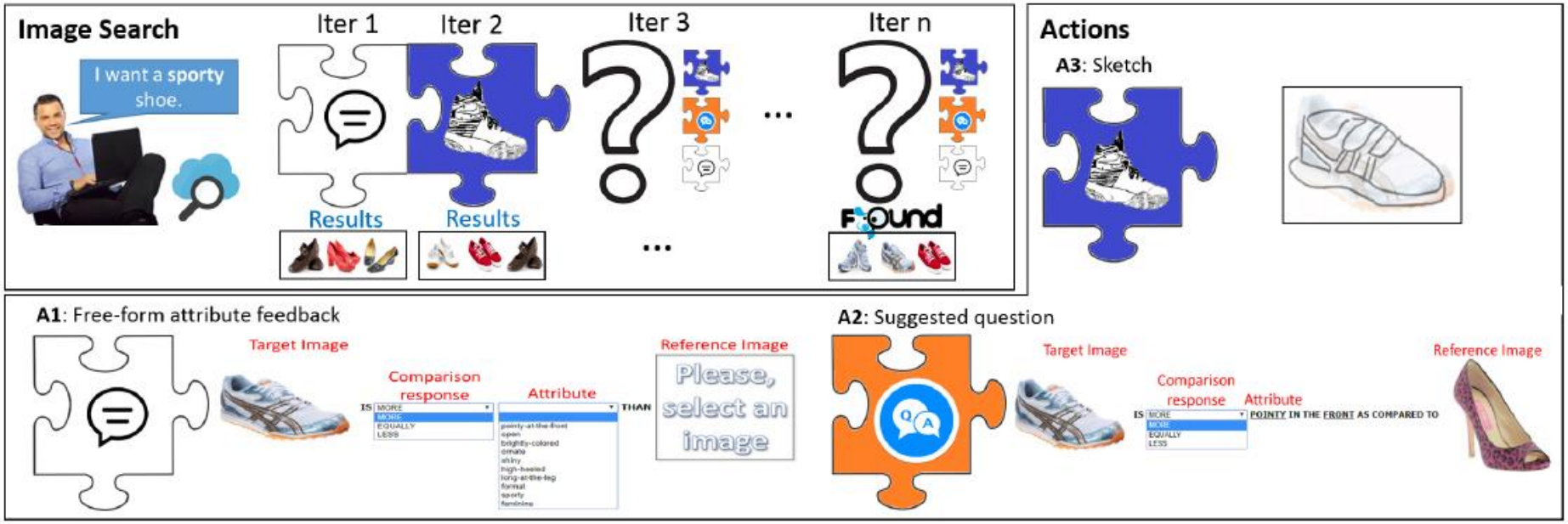
Properties of Reinforcement Learning

- The **agent** is trained for maximizing the **reward**
- The **agent's action** effects its next input
- No supervisor present

Main idea

Overall process of image retrieval

- The user initiates a search with images or simple query

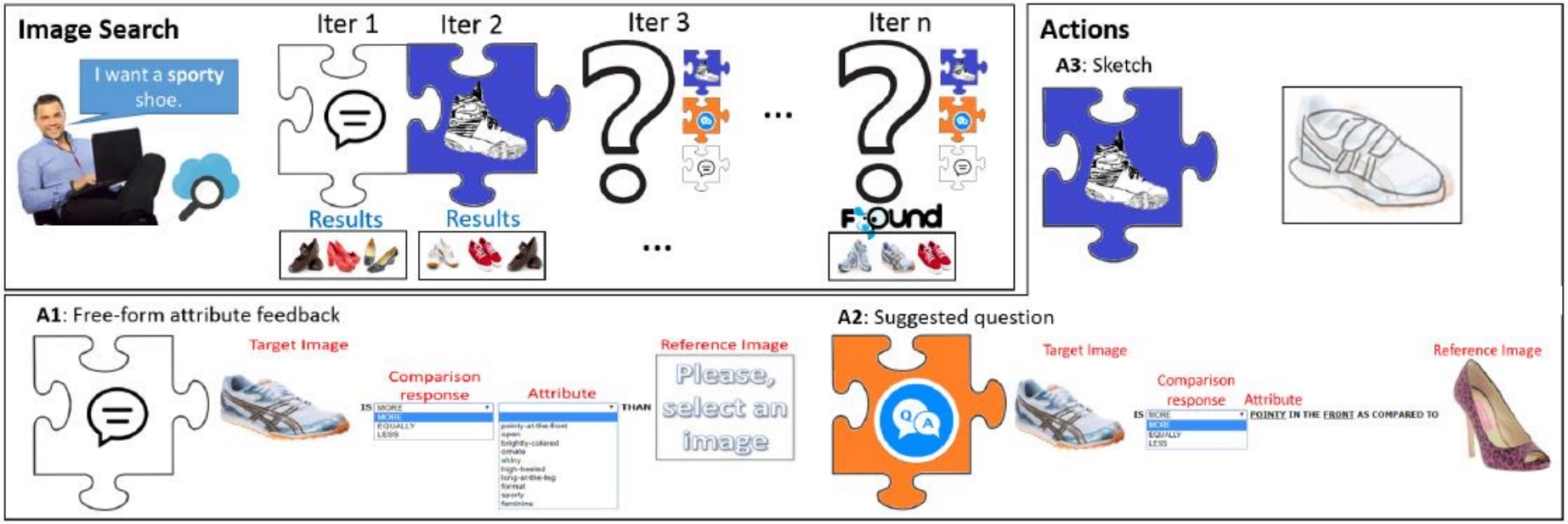


 MURRUGARRA et al., BMVC '18

KAIST

Overall process of image retrieval

- The system chooses the **best feedback method** for retrieving the target image

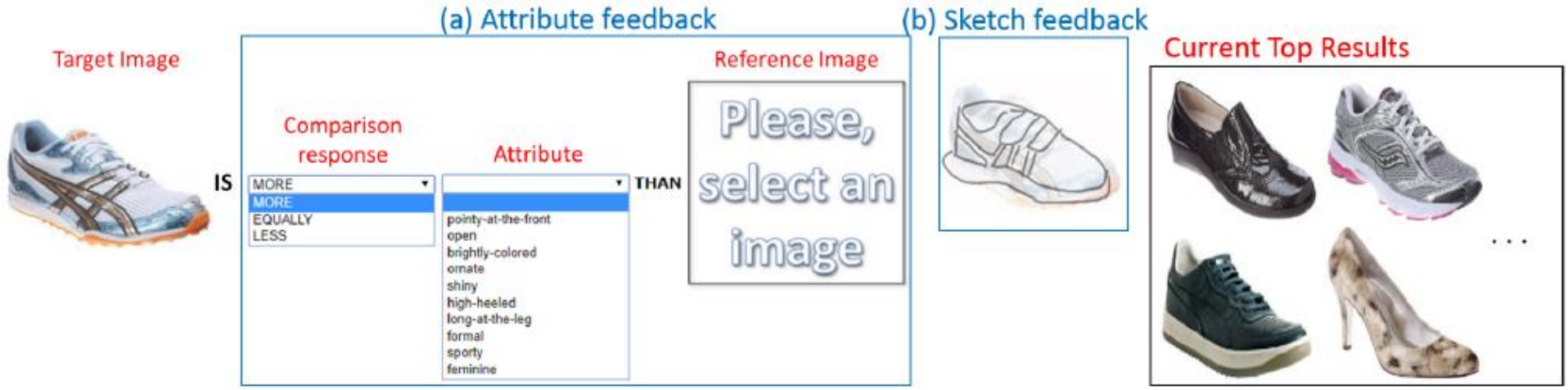


 MURRUGARRA et al., BMVC '18

KAIST

Refining the image retrieval

- Use **(attribute, reference image, comparison response) triplet** or a **sketch**
- The feedback is used to update the next top K results



 MURRUGARRA et al., BMVC '18

KAIST

Refining the image retrieval

- Sketch feedback is converted to photographs by GAN

Figure 6: Sample sketch-to-photo colored images for Pubfig (columns 1-3), Shoes (columns 4-6), and Scenes (columns 7-9). Each column denotes a different category.

 MURRUGARRA et al., BMVC '18

Training the agent with RL

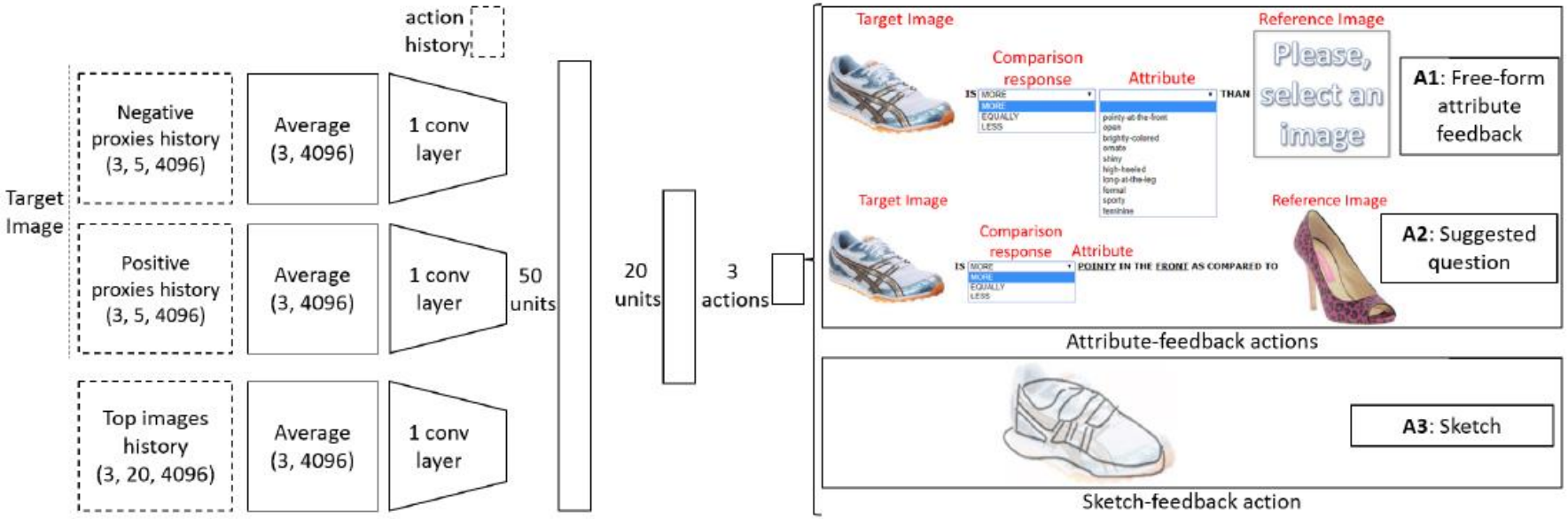
- State – (History of top 1 image, 5 closest neighbor, 5 furthest neighbor, previous action)
- Action – Selecting interaction method

 MURRUGARRA et al., BMVC ‘18

Training the agent with RL

- Reward: Getting bigger reward when the distance to nearest decreases, and distance to furthest increases

$$R(s,s') = sign[d(top\_ims,+prox) - d(top\_ims',+prox)] + sign[d(top\_ims',-prox) - d(top\_ims,-prox)]$$

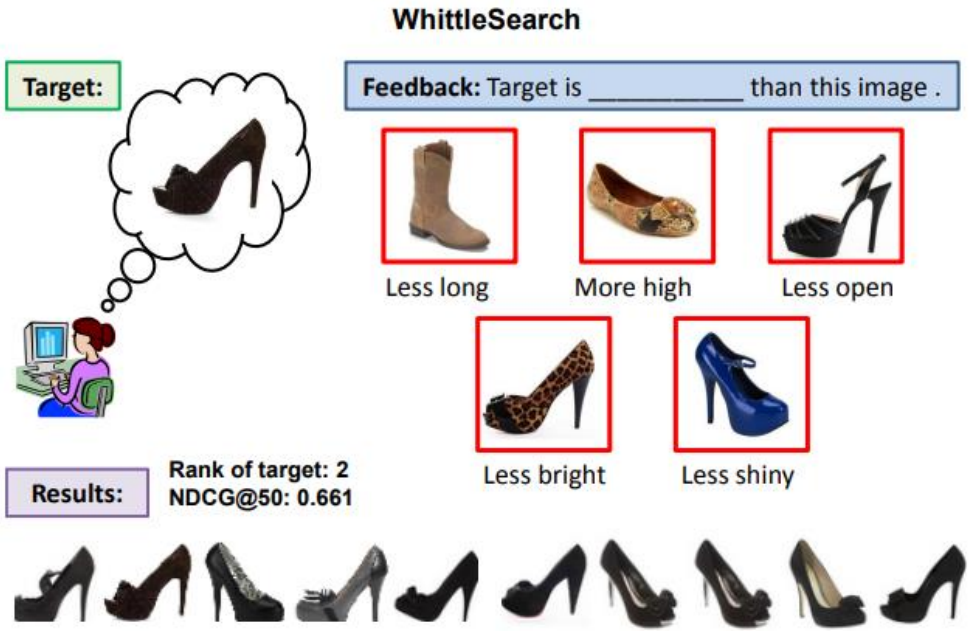


 MURRUGARRA et al., BMVC '18

Experiment & Results

Experiment setup

- Experiment on 3 dataset
  - **Pubfig, Scenes, Shoes**
- Comparing with 3 baselines
  - Whittle search (WS)



 MURRUGARRA et al., BMVC ´18, Kovashka et al., IJCV ´15

KAIST

Experiment setup

- Experiment on 3 dataset
  - **Pubfig, Scenes, Shoes**
- Comparing with 3 baselines
  - Whittle search (WS)
  - Pivot round robin (PRR)

 MURRUGARRA et al., BMVC '18, Kovashka et al., ICCV '13

KAIST

Experiment setup

- Experiment on 3 dataset
  - **Pubfig, Scenes, Shoes**
- Comparing with 3 baselines
  - Whittle search (WS)
  - Pivot round robin (PRR)
  - Sketch retrieval + PRR (SK_PRR)
- Experimented by simulated and real users

Results

- Percentile rank plots for each dataset

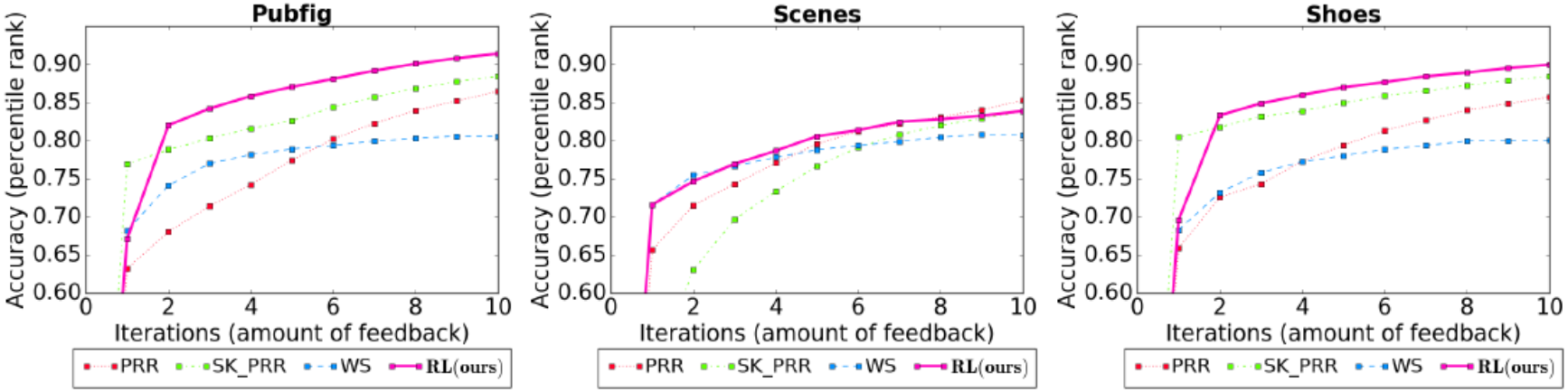


Figure 4: Percentile rank plots for Pubfig, Scenes, and Shoes. Our mixed-initiative RL agent outperforms the other baselines on Pubfig and Shoes, and performs competitively for Scenes.

 MURRUGARRA et al., BMVC '18

KAIST

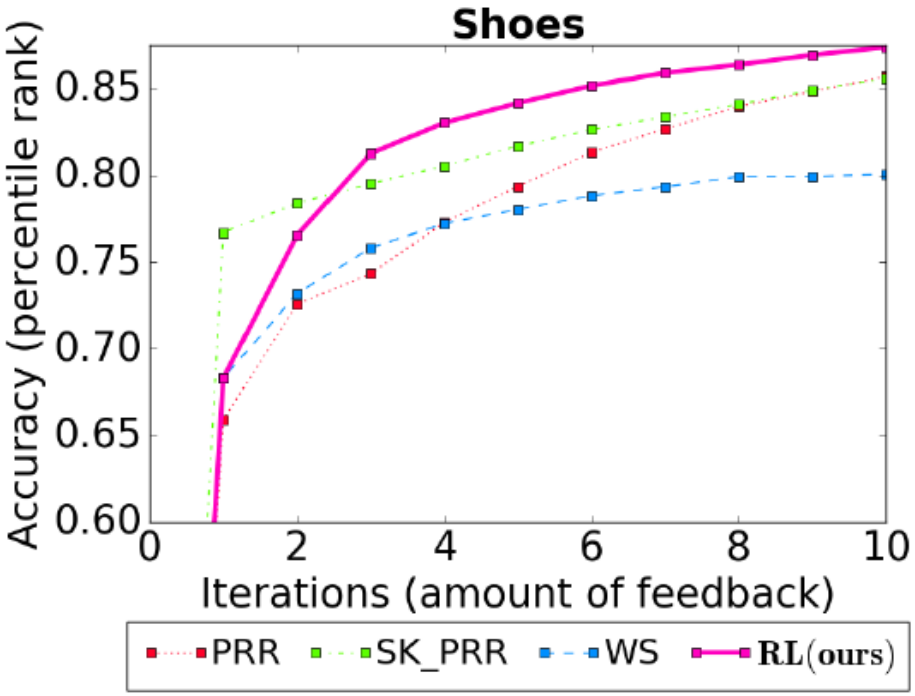
Results

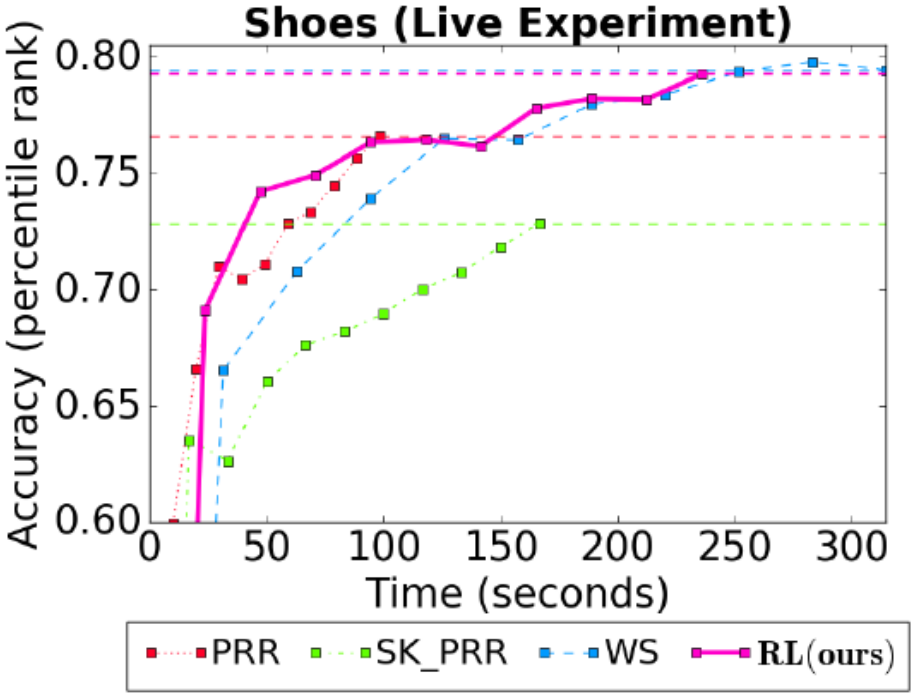
- AUC for percentile rank curves

	PRR [16]	WS [19]	SK_PRR [16, 54]	**RL (ours)**
Pubfig	0.729	0.737	0.789	**0.810**
Scenes	0.741	0.741	0.699	**0.754**
Shoes	0.745	0.731	0.806	**0.810**
avg	0.738	0.736	0.764	**0.791**

Results

- Live experiment result





 MURRUGARRA et al., BMVC ‘18

KAIST

Thank you!

KAIST