WhittleSearch: Image Search with Relative Attribute Feedback

Adriana Kovashka, Devi Parikh and Kristen Grauman

(Supplementary Material)

1 Comparative Qualitative Search Results

We present three qualitative search results for human-generated feedback, in addition to those shown in the paper. Each example shows one search iteration, where the 20 reference images are randomly selected (rather than ones that match a keyword search, as the image examples in the main paper illustrate). For each result, the first figure shows our method and the second figure shows the binary feedback result for the corresponding target image. Note that for our method, "more/less X" (where X is an attribute) means that the target image is more/less X than the reference image which is shown.

Figures 1 and 2 show results for human-generated relative attribute and binary feedback, respectively, when both methods are used to target the same "mental image" of a shoe shown in the top left bubble. The top right grid of 20 images are the reference images displayed to the user, and those outlined and annotated with constraints are the ones chosen by the user to give feedback. The bottom row of images in either figure shows the top-ranked images after integrating the user's feedback into the scoring function, revealing the two methods' respective performance. We see that while both methods retrieve high-heeled shoes, only our method retrieves images that are as "open" as the target image. This is because using the proposed approach, the user was able to comment explicitly on the desired openness property.

Figure 1: Shoes search example, using human-generated relative attribute feedback.

Figure 2: Shoes search example, using human-generated binary feedback.

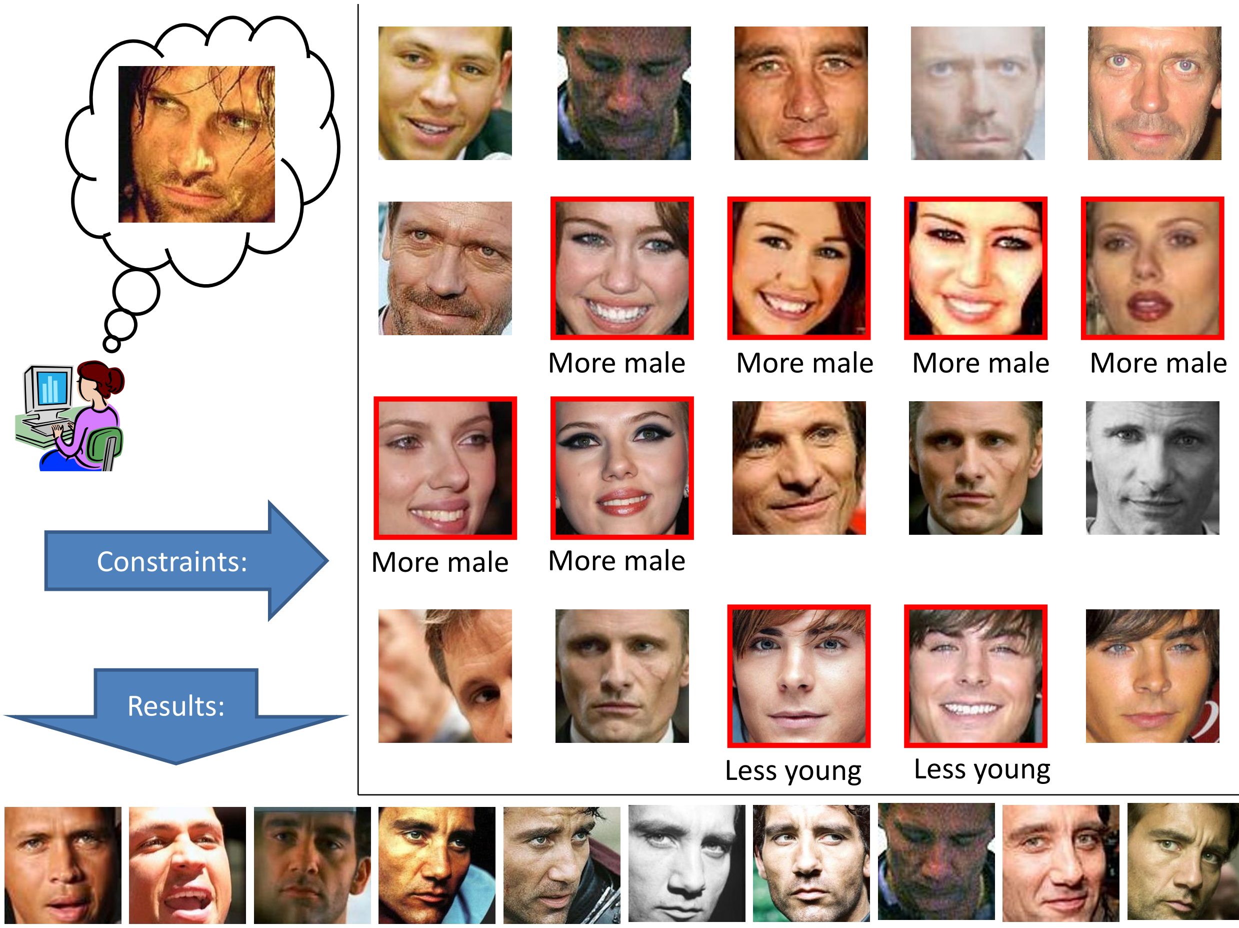

Figure 3: PubFig search example, using human-generated relative attribute feedback.

Figures 3 and 4 show a failure case for our method on the PubFig data, using the same format as described above. The failure is mostly due to the imprecise user-specified constraints, which are not sufficient to distinguish the person in the target image from other "older" males. In contrast, the binary feedback baseline (Figure 4) successfully learns a "Viggo Mortensen" classifier, since the reference images happen to contain other images of Viggo Mortensen. Still, the binary feedback's results are not exactly the envisioned exemplar of Viggo shown in the top left bubble, even though it is present in the database pool. Whether this result would be sufficient to a user depends on whether he/she aims to perform a "category-level" search (where the target is a person class), or truly wants an image closely aligned with the envisioned target.[1]

[1]The PubFig dataset contains a number of near-duplicates of the top result image in Figure 4, and the binary feedback baseline ranks many of these duplicates high, so we omit some of them from the shown results and instead show some of the other images ranked in the top 20. Note that the top-ranked image and its duplicates present a rarity, in that it is uncommon for an image to have this many duplicates in the PubFig dataset.

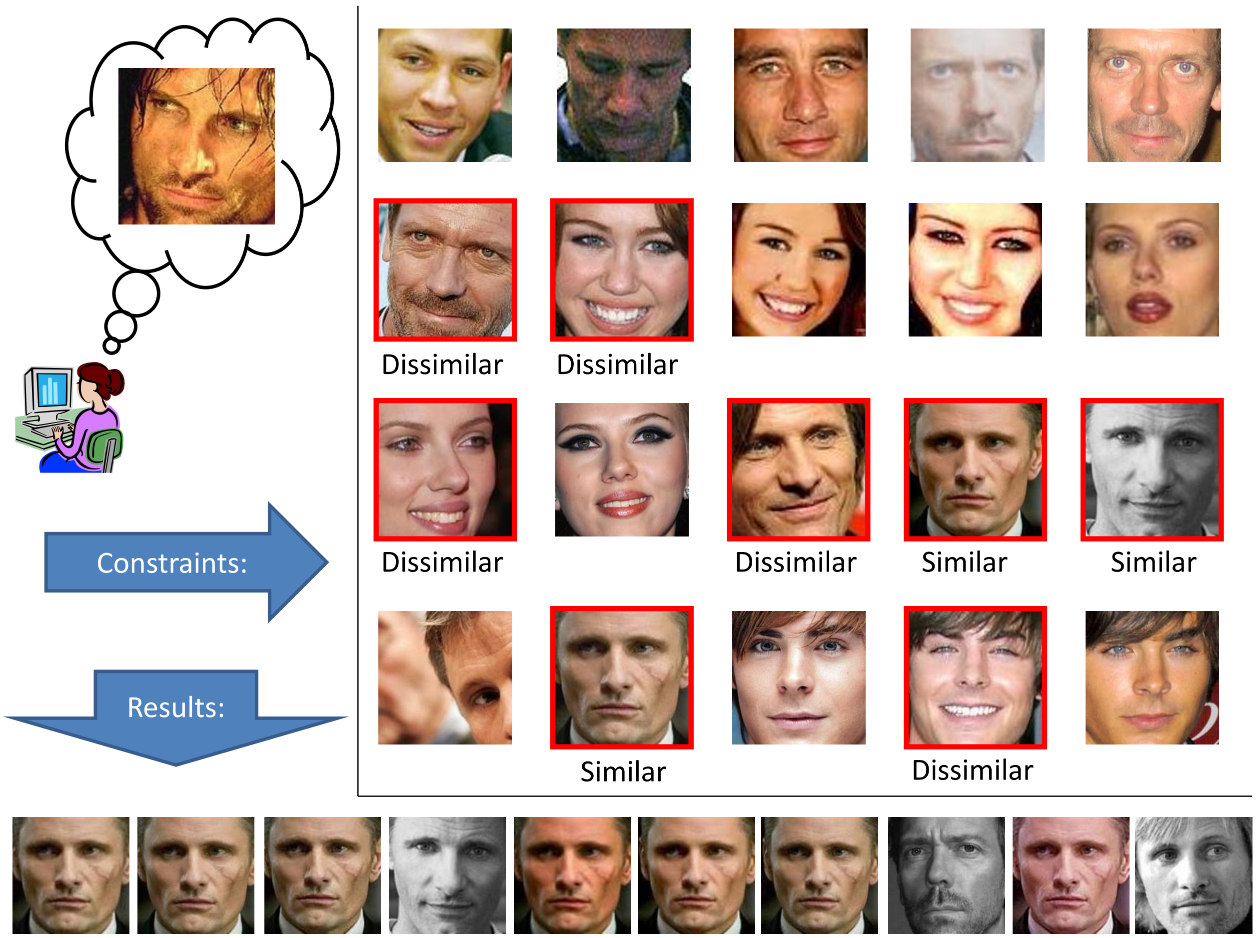

Figure 4: PubFig search example, using human-generated binary feedback.

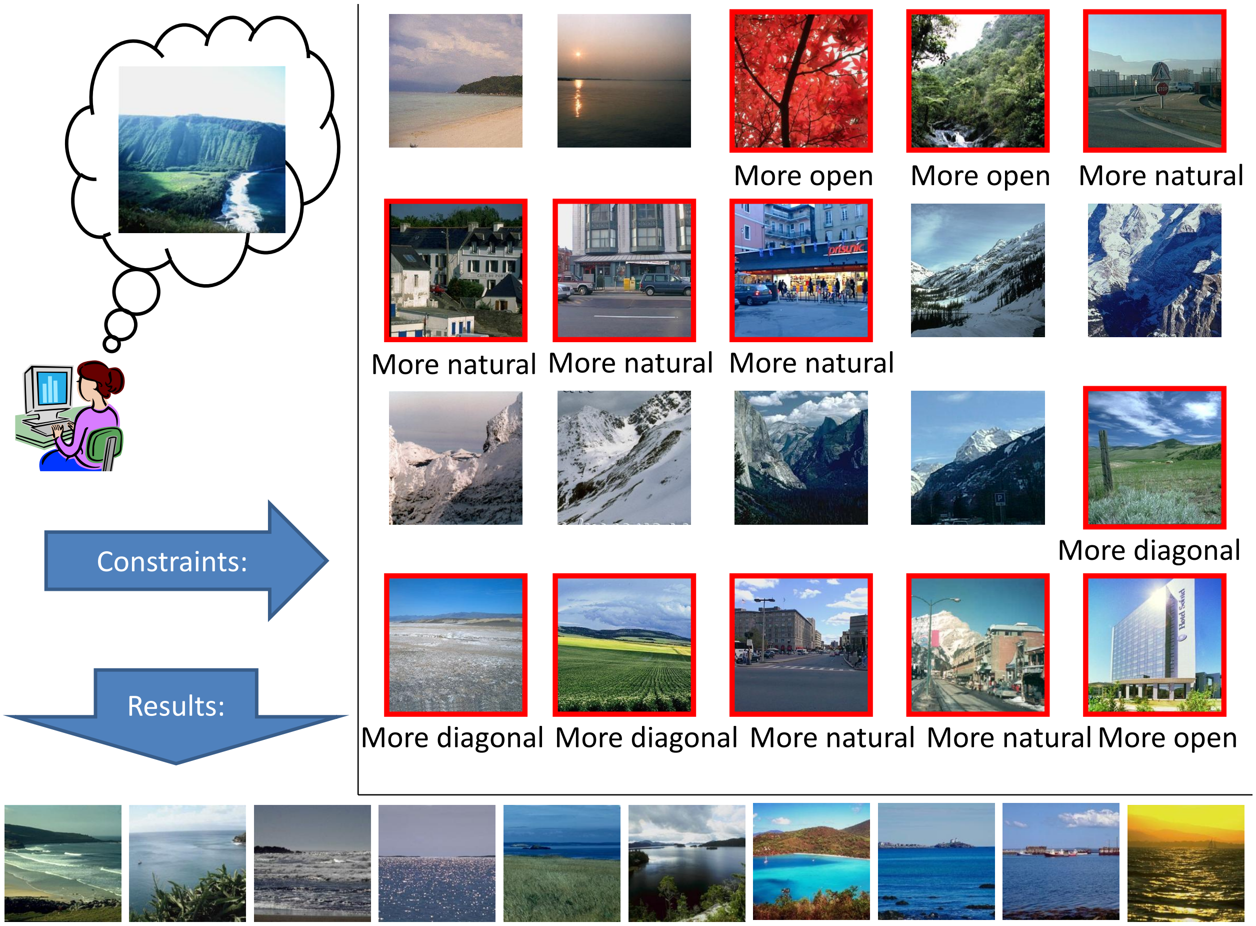

Figure 5: OSR search example, using human-generated relative attribute feedback.

Figures 5 and 6 show an interesting example of a target image that is hard to describe in words and likely has few very similar images in the database. However, through our relative attribute constraints, we are able to retrieve better matches (Figure 5) than the binary feedback baseline produces (Figure 6). A main issue for the baseline in this case is the lack of similar images among the reference images which the humans can use to define positives, as well as the presence of only one positive which results in learning a “tree scene” classifier.

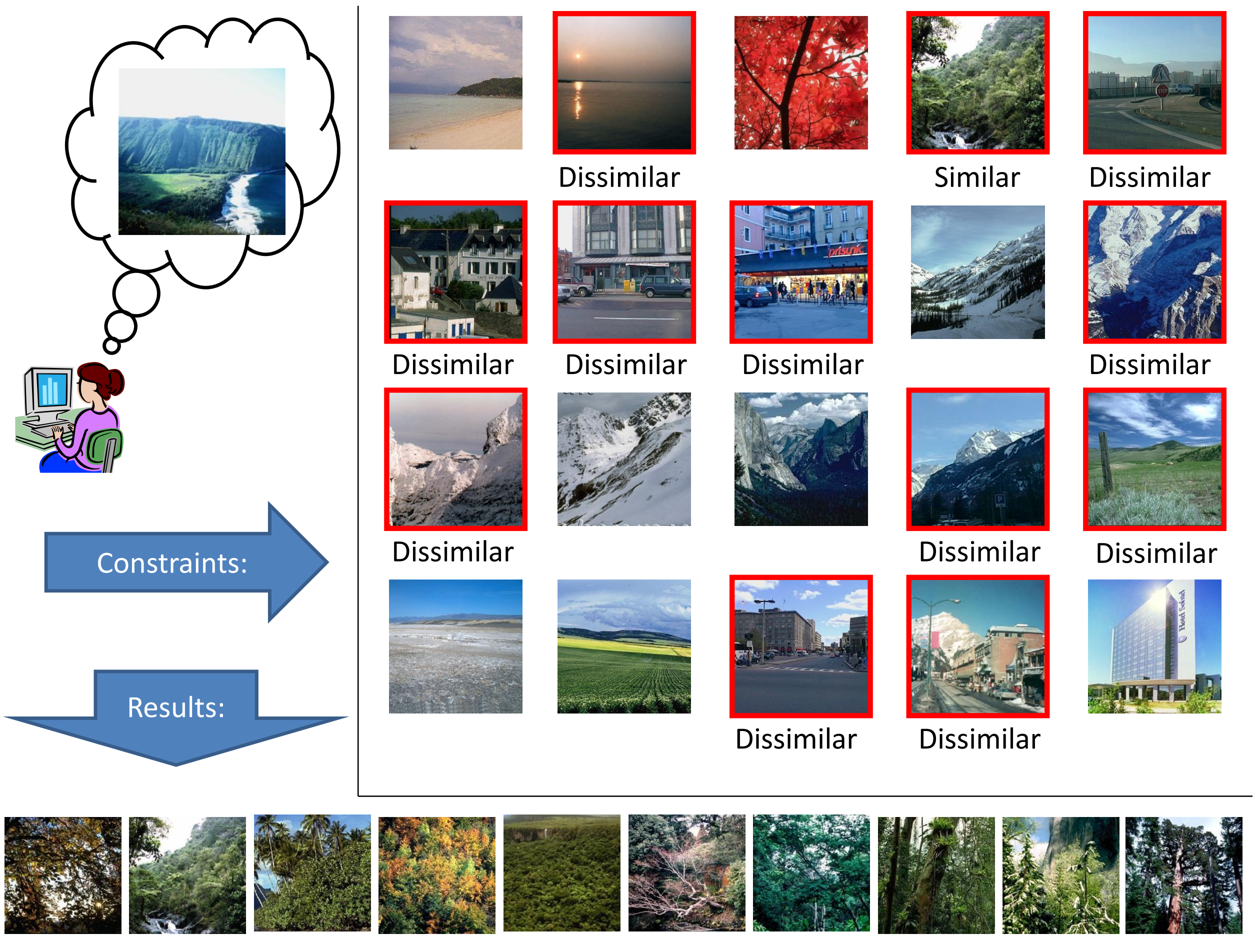

Figure 6: OSR search example, using human-generated binary feedback.

2 Class-level Attribute Orderings on Shoes Dataset

Table 1 shows the relative attribute orderings we defined between the different types of shoes for the Shoes dataset. Each row corresponds to an attribute. The numbers in each row indicate how much the class to which the number corresponds has the attribute, with 10 denoting "has it the most" and 1 denoting "has it the least". These orderings allow us to train class-level relative attributes in Section 4.3 of the main paper, for comparison with the proposed instance-level training procedure.

Attribute/Class	Athletic	Boots	Clogs	Flats	Heels	Pumps	Rain Boots	Sneakers	Stiletto	Wedding
Pointy at the front	2	6	3	5	10	9	4	1	8	7
Open	3	2	8	5	7	6	1	4	9	10
Bright in color	6	1	2	8	4	3	10	7	9	5
Covered w/ ornaments	4	9	6	5	8	7	1	3	10	2
Shiny	2	9	4	3	6	5	8	1	10	7
High at the heel	4	6	5	1	9	8	3	2	10	7
Long on the leg	7	9	2	3	6	5	10	8	4	1
Formal	3	6	4	7	9	8	1	2	5	10
Sporty	10	5	6	7	4	3	8	9	1	2
Feminine	1	6	4	5	10	9	3	2	8	7

Table 1: Orderings of classes for the attributes in the Shoes dataset.