FIRE-CIR: Fine-grained Reasoning for Composed Fashion Image Retrieval

Supplementary Material

1. Additional FIRE-CIR qualitative examples

To further illustrate the reasoning process, we present additional qualitative examples of FIRE-CIR on the Fashion IQ dataset ("dress" subset for Figure 1 and Figure 2, and "shirt" subset for Figure 3 and Figure 4).

For each example, the reference image and modification text are in the top-left corner. FIRE-CIR decomposes the modification text into a set of visual questions, listed on the left. These questions are then applied to the top-6 images retrieved by FashionBLIP-2 (from the first image on the top left to the sixth on the top right), and to the ground-truth target image (top-right corner). Each checkmark (resp. cross) corresponds to a predicted answer compatible (resp. incompatible) with the modification text. The probability of the answer being compatible with the text is displayed below the checkmark (resp. cross). Then, all answer probabilities are averaged for each candidate image to compute the VQA score, which is used to compute the final rank of each candidate image. The final rank is indicated at the bottom of each example. We also specify the rank difference with the FashionBLIP-2 results: a negative difference means that the candidate image gets a higher rank (so it is more relevant), while a positive difference means that the candidate image is further among the retrieved results (so it is less relevant).

Figure 1. In this example, FIRE-CIR is able to accurately identify the pattern in the dresses and re-ranks the retrieved results according to their resemblance with what is described in the modification text.

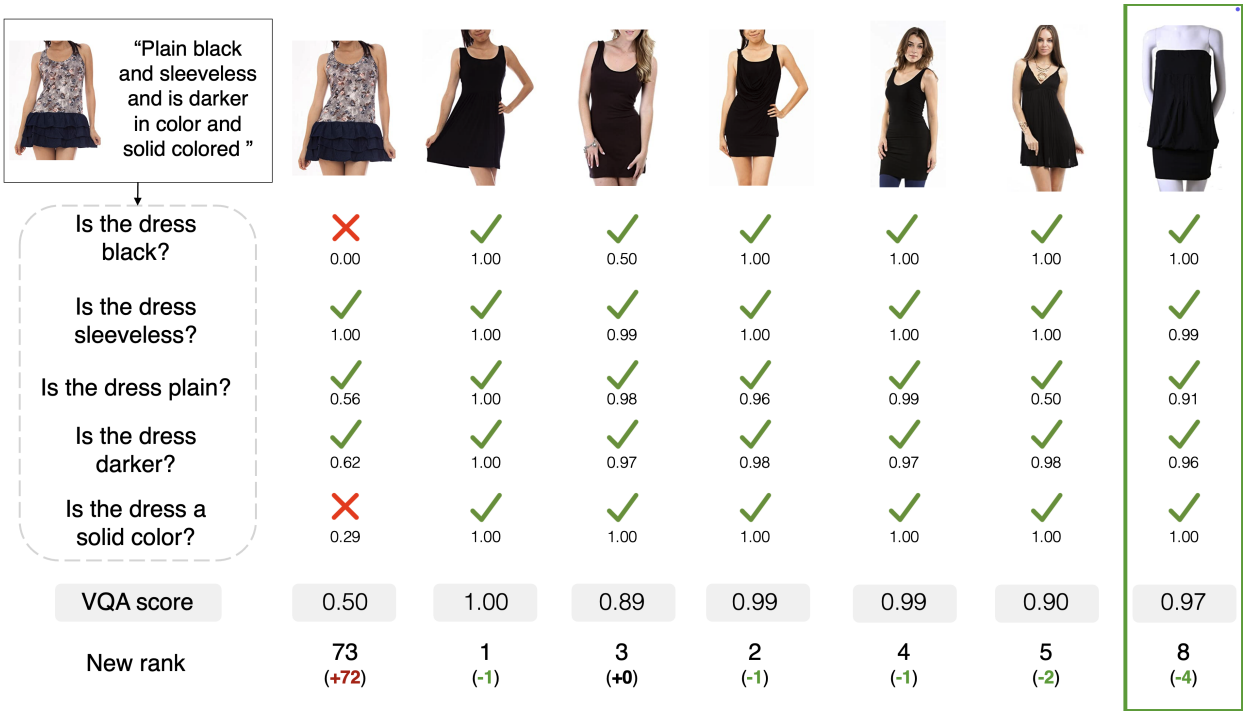

Figure 2. While FashionBLIP-2 ranks highly the reference image as some of its visual aspects remain compatible with the CIR query, FIRE-CIR correctly identifies that this dress does not have a solid black color, and thus re-ranks it outside of the top-50 results. The ground-truth target image is at the 8th rank, but the 7 other top-ranked images are also compatible with the given query.

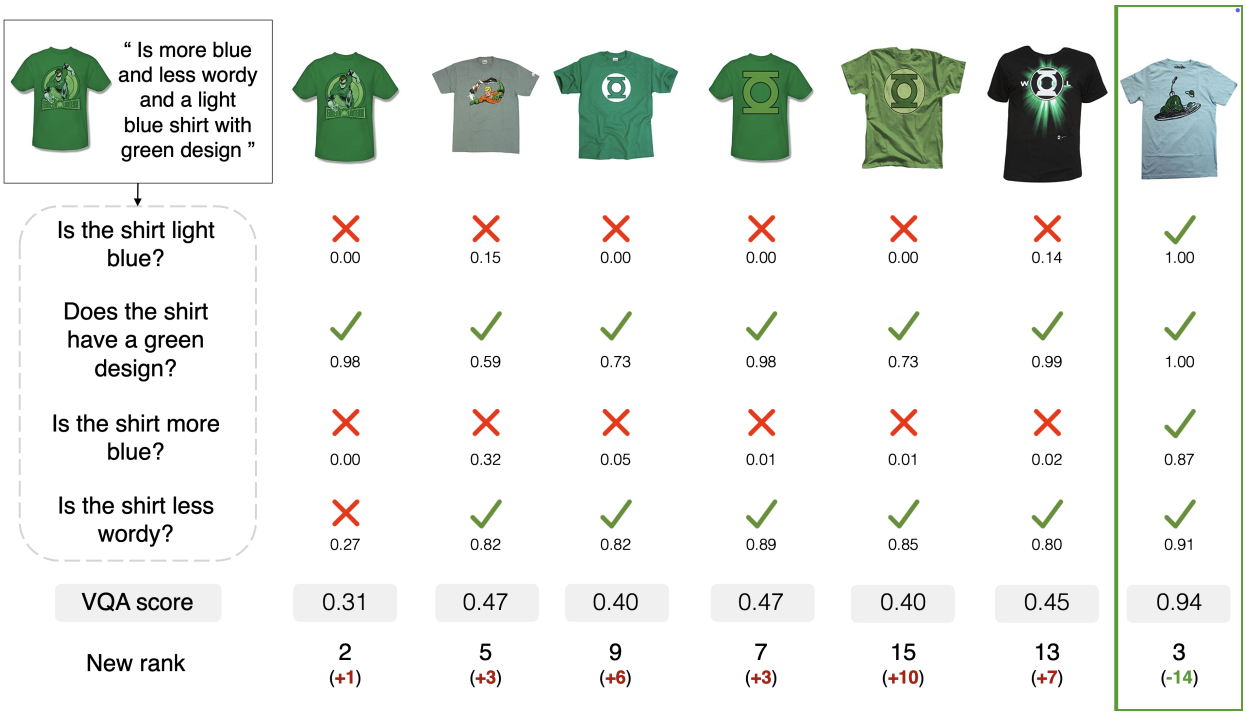

Figure 3. Similarly, visual similarity and the green design contribute to having candidate images similar to the reference one in the top-retrieved results. However, FIRE-CIR accurately promotes the ground-truth image, as its color is more coherent with what is written in the modification text.

Figure 4. Contrary to FashionBLIP-2 which focuses specifically on the "workout muscle" characteristic, FIRE-CIR gives more importance to all the features, including the orange color.