Image Retrieval with Mixed Initiative and Multimodal Feedback
(Supplemental Material)

Nils Murrugarra-Llerena
nineil@cs.pitt.edu

Adriana Kovashka
kovashka@cs.pitt.edu

Department of Computer Science
University of Pittsburgh
Pittsburgh, PA, USA

As supplemental material, we provide qualitative results for sketch-to-photo coloring and retrieval, which is one crucial step in our approach. We also show examples of sketches that our live users drew.

1 Edge maps and conversions to photos

We complement our qualitative results on generated sketch-to-photo pictures (Fig. 1). We observe that the most realistic ones correspond to Pubfig, then Shoes, and finally Scenes. We observe again the poor quality of Scenes generated images. However, our GAN associates brown color to coast images (panels 1 and 2 in row 6, from Fig. 1). Similarly, it learns green color for forest images (panels 5 and 6 in row 6, from Fig. 1). Even apart from the generations' quality, edge maps from Scenes do not provide as much detail as edge maps for Faces and Shoes. For example, only the exterior surface of buildings was present in the edge map (see last two panels in row 5). High-level edge maps also can remove crucial objects in the scene, that can not be colored. For example, some trees were removed in (row 5, column 6), which hampers coloring.

2 Sketches from live users

We examine provided sketches from our live users in Fig. 2. We observe that many of them do a good job. For example, in (row 1, column 4), the sketch has finer details such as the flower ornaments of the flat shoe. Similarly, for (row 3, column 1), the boot was drawn with laces in its top as in its middle. Finally, a sneaker sketch (row 3, column 2) contains shoelaces and details at its bottom part.
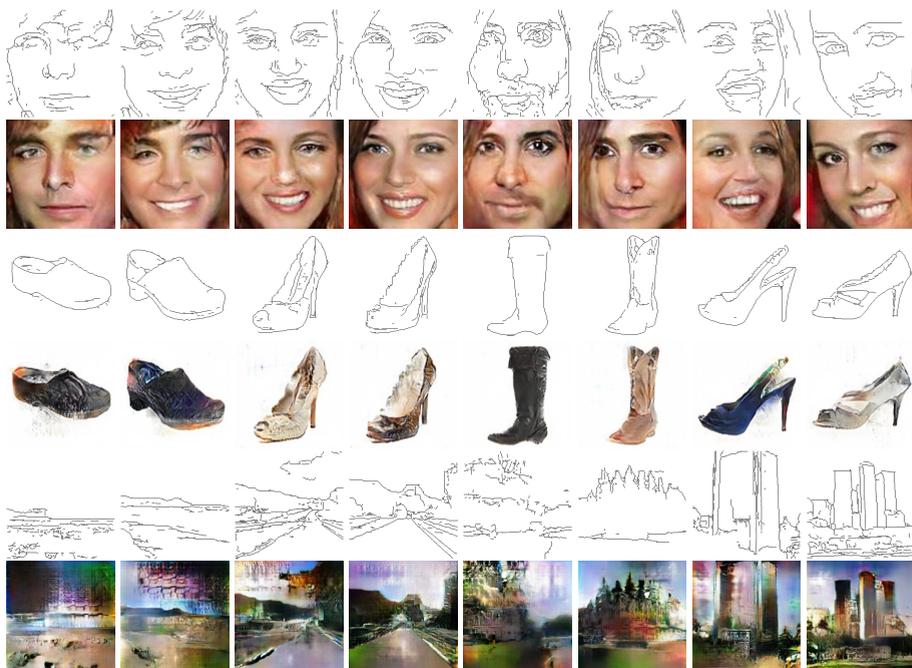
Figure 1: Sample sketch-to-photo colored images for Pubfig (rows 1-2), Shoes (rows 3-4), and Scenes (rows 5-6). Each pair of images denotes the same class category. For each dataset, the first row shows the edge maps, and the second row shows the colored picture.

Figure 2: Sketches provided by annotators from Amazon Mechanical Turk and university students for our live experiment. Rows 1 and 3 are user sketches, and rows 2 and 4 are their correspondent target images.