A. Additional Results

Figs. 1-6 depict additional sketch-to-photo generation results with sketches from QMUL-ShoeV2 [1, 6], QMUL-ChairV2 [1, 6], and Handbag [7] datasets. A common observation seen in Figs. 1-6 is that, although photorealistic, the output quality is comparatively lower in ChairV2 and Handbag datasets than that in ShoeV2 due to a higher degree of sketch abstraction and a lower number of training pairs available in the former two datasets.

Figure 1. Results generated by the proposed method with sketches from the QMUL-ShoeV2 [1, 6] dataset.
Figure 2. Results generated by the proposed method with sketches from the QMUL-ShoeV2 [1, 6] dataset.
Figure 3. Results generated by the proposed method with sketches from the QMUL-ChairV2 [1, 6] dataset.
Figure 4. Results generated by the proposed method with sketches from the QMUL-ChairV2 [1, 6] dataset.
Figure 5. Results generated by the proposed method with sketches from the Handbag [7] dataset.
Figure 6. Results generated by the proposed method with sketches from the Handbag [7] dataset.
B. Details on Human Study

Fig. 7 and Fig. 8 depict the login and scoring screen of the interface used to collect the MOS [3] values. Upon login, a user first selects a class (i.e., Shoe, Chair, or Handbag) and draws a sketch of that class. Next, upon clicking on the “Generate” button, the system displays corresponding photo translations produced by our proposed method along with every other competing framework. The participant rates every generated photo and clicks on “Submit & Next” to continue. We anonymise the names of the competing methods (e.g., pix2pix [4], MUNIT [2], etc.) to prevent the ratings from being influenced by the participant’s past knowledge. For brevity and ease of the participants, following [3], we sub-divide the 1→5 (bad→excellent) MOS levels into a nine-point discreet scale with the possible ratings \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}. For each method, we compute the final MOS [3] value by taking the mean (µ) and variance (σ) of all individual MOS responses.

![Figure 7. Login screen of our human study interface](image)

![Figure 8. Scoring screen of our human study interface](image)
C. Intermediate Photo Generation

Leveraging StyleGAN’s [5] smooth latent space, our method allows us to generate realistic transitional photos given an initial and a final sketch. We achieve this via simple arithmetic operations (e.g., interpolation) between the predicted latent codes of the two sketches. Fig. 9 and 10 shows how given Sketch A and Sketch B, our method can generate plausible intermediate photos.

Figure 9. Generating transitional shoe photos between Sketch A and B.

Figure 10. Generating transitional handbag photos between Sketch A and B.

References