Generating Synthetic Data via Augmentations for Improved Facial Resemblance in DreamBooth and InstantID

Supplementary Material

9. FaceDistance of InstantID generations

Our experiments with InstantID reveal that using a single reference image produces high variance in output quality with inconsistent facial similarity. (Figure 5) When using multiple reference images, results significantly improve facial representation consistency (Figure 6). Analysis of the *Vacation-Anna* dataset demonstrates that while the visual improvement between using 1 and 2 reference images is substantial, the differences between using 2, 4, or 8 images are subtle and difficult to distinguish without direct side-by-side comparison. The FaceDistance metric confirms this, showing that as more reference images are used, the distribution of distances shifts leftward, indicating better overall similarity.
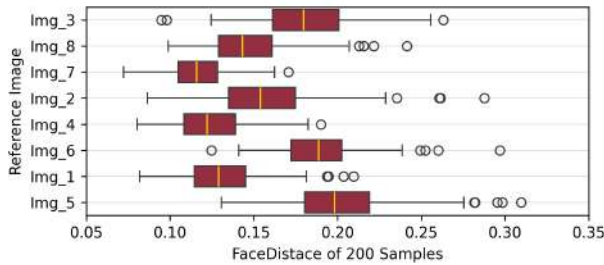

Figure 5. Distribution of facial similarity metrics for 8 different input subjects using single reference images, with the y-axis sorted by FaceDistance (Img 3 showing lowest inter-image distance, Img 5 showing highest).
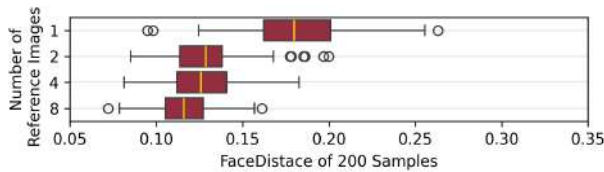

Figure 6. Comparison of facial similarity performance when using k reference images, demonstrating the leftward shift in FaceDistance distribution as k increases from 1 to 8 images.

10. Motivation to Seek Alternatives to InstantID

Although InstantID produces visually appealing images, its output variability for identical configurations is limited (see Figure 8). Our survey reveals that users are seeking greater diversity in images while maintaining a natural appearance, free from a photoshopped aesthetic. In contrast, utilizing Dreambooth for image generation circumvents this issue, making it the more favorable option (Figure 7).

Figure 7. Dreambooth generated images with more variability *(Vacation-Anna)*

11. Comparative Survey Analysis of Portrait Generation Methods

Our analysis reveals that generated portraits perform similarly across individual aspects, likely due to the SD model employed (Figure 9). Good subject datasets for Dreambooth yield more natural results, though user preferences trend toward Photoshop-edited appearances. Further investigation shows that "good" datasets correlate with higher employee agreement on preferred methods, with approximately 4% more participants favoring standardized InstantID portraits for their professional, photoshopped aesthetic (Figure 10). When assessing facial similarity between real and generated images, Dreambooth consistently outperformed InstantID regardless of dataset quality, with fewer participants identifying Dreambooth generations as different individuals (Figure 11).

White-collar workers demonstrated limited ability to distinguish AI-generated images, often searching for well-

Figure 8. InstantID generated images with better facial similarity but with less variability *(Party-Ryan)*

known AI flaws absent from our high-quality generations. Among the subset of participants (n = 77) who reported noticing AI-generated images in daily life, our generations integrated seamlessly with non-generated studio images, with only the "Real (Daniel)" image—featuring strong contrast and sharp features—dividing opinion nearly equally (Figure 12).
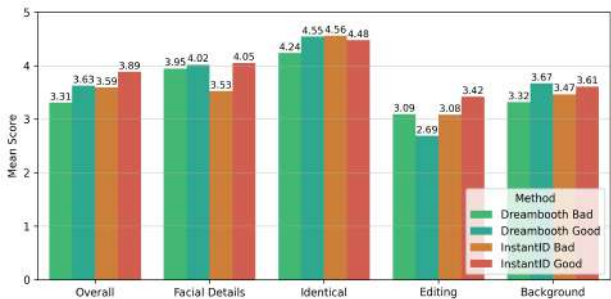

Figure 9. Mean Scores by Category for Each Method. Users were asked to rate samples from each method on 5 Categories, 5 being the highest value.

12. Survey Questionare

Here is the full questionnaire for the survey. The answer options are in *italic*. Titles and other guiding instructions are in **bold**. We included descriptions to the images.

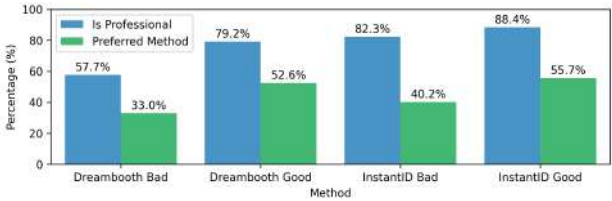**Title: Survey on Studio Portraits**

Short introduction
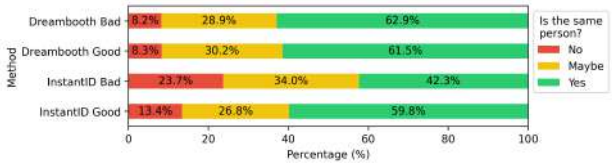

Figure 10. Percentage of Participant Choices.

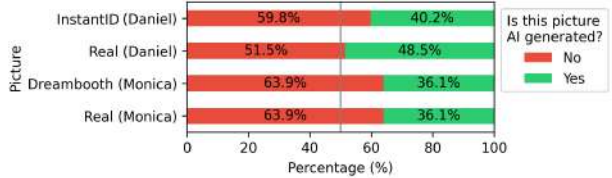Figure 11. Percentage of Similarity Ratings per Method.

Figure 12. Percentage of No/Yes Answers per Picture. $(n = 77)$

1. **Rate the skill level of the photographer. (This is repeated for all four subjects.) Each subject has four generated images arranged in a grid. A short description of their professional background is included (e.g., marketer, researcher, nurse).**
   (a) How would you rate the overall quality? *[Really Bad] 1 - 5 [Really Good]*
   (b) Are the facial details clear and well-defined? *[No] 1 - 5 [Yes]*
   (c) How identical is the person in these pictures? *[Completely Different] 1 - 5 [Exactly The Same]*
   (d) How much editing, if any, is present in this photo? *[No Editing] 1 - 5 [Heavily Edited]*
   (e) Rate the quality of the background in the headshot. *[Poor] 1 - 5 [Excelent]*
   (f) Would you expect to see these photos in a professional context, such as on LinkedIn or company websites? *[Yes/No]*
   (g) Is there anything you don't like about these pictures? *[Text Answer]*
   (h) Is there anything you particularly like about these pictures? *[Text Answer]*
2. Which photographers were good? If you've liked at least 1 image, click the square near it. *[Four image grids from above]*
3. **Similarity of Real-Life Pictures and Portraits. In this section, you are asked to compare how similar**

the person in the everyday pictures on the left is to the person in the portraits on the right.

How would you rate the similarity of the person? *[Not the same] 1-3 [The Same]*

4. **Detection of the use of AI.** **In this chapter, we might have used artificial intelligence (AI) in a way. Don't go back to previous chapters and change your answers. This is important for our research.**

Please select photos, if any, that you think AI was used. *[Four pictures of the subjects]*

5. What affected your choice? *[Text Answer]*

6. **Your familiarity with photography**

How often do you take photos of yourself or others in your daily life? *[Daily or Weekly, Monthly, Every few Months, Yearly or Never]*

7. Do you use any software to edit your photos? *[Yes/No]*

8. Do you spot AI (artificial intelligence) generated pictures in your daily life? *[Yes/No]*

9. Have you ever used AI to create pictures? *[Yes/No]*

References

[1] ComfyUI_InstantID, 2025. [Online; accessed 30. Mar. 2025]. 3

[2] SG161222/RealVisXL_V4.0 · Hugging Face, 2025. [Online; accessed 30. Mar. 2025]. 3

[3] TheBloke/dolphin-2.2.1-mistral-7B-GGUF · Hugging Face, 2025. [Online; accessed 30. Mar. 2025]. 3

[4] sd-scripts, 2025. [Online; accessed 30. Mar. 2025]. 3

[5] Category:Patterns - Wikimedia Commons, 2025. [Online; accessed 31. Mar. 2025]. 3

[6] Aditya Agarwal, Bipasha Sen, Rudrabha Mukhopadhyay, Vinay Namboodiri, and C. V. Jawahar. FaceOff: A Video-to-Video Face Swapping System. *arXiv*, 2022. 2

[7] Ting-Yu Chang and Seretsi Khabane Lekena. GlamTry: Advancing Virtual Try-On for High-End Accessories. *arXiv*, 2024. 2

[8] Khawar Islam and Naveed Akhtar. Context-guided Responsible Data Augmentation with Diffusion Models. *arXiv*, 2025. 2

[9] Khawar Islam, Muhammad Zaigham Zaheer, Arif Mahmood, and Karthik Nandakumar. DiffuseMix: Label-Preserving Data Augmentation with Diffusion Models. *arXiv*, 2024. 2

[10] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. FaceShifter: Towards High Fidelity And Occlusion Aware Face Swapping. *arXiv*, 2019. 2

[11] Jun Hao Liew, Hanshu Yan, Daquan Zhou, and Jiashi Feng. MagicMix: Semantic Mixing with Diffusion Models. *arXiv*, 2022. 2

[12] Alex Nichol and Prafulla Dhariwal. Improved Denoising Diffusion Probabilistic Models. *arXiv*, 2021. 2

[13] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. *arXiv*, 2023. 3

[14] Xuebin Qin, Zichen Zhang, Chenyang Huang, and Masood Dehghan. U^2-Net: Going Deeper with Nested U-Structure for Salient Object Detection. *arXiv*, 2020. 3

[15] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis With Latent Diffusion Models, 2022. [Online; accessed 31. Mar. 2025]. 2

[16] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. *arXiv*, 2022. 2

[17] Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. *arXiv*, 2015. 2, 4

[18] Sefik Serengil and Alper Özpınar. A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules. *Bilişim Teknolojileri Dergisi*, 17(2):95–107, 2024. 2

[19] Brandon Trabucco, Kyle Doherty, Max Gurinas, and Ruslan Salakhutdinov. Effective Data Augmentation With Diffusion Models. *arXiv*, 2023. 2

[20] Qixun Wang, Xu Bai, Zekui Qin, Anthony Chen, Huaxia Li, Xu Tang, and Yao Hu. InstantID: Zero-shot Identity-Preserving Generation in Seconds. *arXiv*, 2024. 2, 3, 6

[21] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, and Chao Dong. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *arXiv*, 2018. 3

[22] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *arXiv*, 2016. 2