

## Checklist

1. For all authors...
  - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? [Yes] We clarify our contributions relative to prior work in Section 1.
  - (b) Did you describe the limitations of your work? [Yes] As we introduce assumptions and our algorithm in Sections 2 and 3, we highlight what is required in order to use our approach. We summarize limitations explicitly in Section 5.
  - (c) Did you discuss any potential negative societal impacts of your work? [Yes] We do so explicitly in Section 5.
  - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
  - (a) Did you state the full set of assumptions of all theoretical results? [Yes] We clarify our main assumption and definitions in Section 2, and each theoretical statement refers back to the required assumptions.
  - (b) Did you include complete proofs of all theoretical results? [Yes] Appendix G contains proofs of all our results.
3. If you ran experiments...
  - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We include a link to the code on the first page of the paper.
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] We specify all of these details in Appendix E for the CelebA experiment, and details of the laboratory testing synthetic experiment are described in Appendix A
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] In our main experiment, we re-run for 100 random seeds, and report results that capture the resulting variation in Figure 5.
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] We give the computation time for our main experiment in Section 4.2, with details of resources in Appendix E. The simpler synthetic examples (e.g., laboratory testing) are run on a personal laptop, not requiring extensive computation.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
  - (a) If your work uses existing assets, did you cite the creators? [Yes] We make use of the CausalGAN code released by Kocaoglu et al. [2018], whom we cite. We also provide a citation for the CelebA dataset.
  - (b) Did you mention the license of the assets? [Yes] See Appendix E
  - (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We include a link to our code, but do not include any other assets.
  - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? [Yes] The CelebA dataset is collected from publicly available images of celebrities on the internet, and we use it as an existing (and popular) research dataset. However, it is not clear that the individuals in question have provided consent, and we make note of this in Appendix E.
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See Appendix E.
5. If you used crowdsourcing or conducted research with human subjects...
  - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
  - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]