Invertible generative models for inverse problems: mitigating representation error and dataset bias

1. Appendix

1.1. Proofs

We now proceed with the proof of Theorem 1. Throughout the proof, we use $\stackrel{d}{=}$ to denote equality in distribution.

Proof of Theorem 1. We first quantify the expected error over the distribution p_G. Observe that since $x_0 \sim p_G = \mathcal{N}(0, GG^{\mathrm{T}})$, it can be written as $x_0 = Gz_0$ where $z_0 \sim \mathcal{N}(0, I_n)$. Then

$$\mathbb{E}_{x_0 \sim p_G} \|\hat{x} - x_0\|^2 = \mathbb{E}_{z_0 \sim \mathcal{N}(0, I_n)} \|G\hat{z} - Gz_0\|^2$$

where $\hat{z}$ is the maximum likelihood estimator in latent space:

$$\hat{z} := \underset{z \in \mathbb{R}^n}{\arg\min} \ \frac{1}{2}\|z\|^2 \text{ s.t. } AGz = AGz_0.$$

Since A has i.i.d. $\mathcal{N}(0,1)$ entries, AG has full rank with probability 1 so $\hat{z}$ is given explicitly by

$$\hat{z} = G^{\mathrm{T}}A^{\mathrm{T}}(AGG^{\mathrm{T}}A^{\mathrm{T}})^{-1}AGz_0 =: \mathcal{P}_{G^{\mathrm{T}}A^{\mathrm{T}}}\, z_0$$

where $\mathcal{P}_{G^{\mathrm{T}}A^{\mathrm{T}}}$ is the orthogonal projection onto the range of $G^{\mathrm{T}}A^{\mathrm{T}}$. Thus

$$\mathbb{E}_{z_0 \sim \mathcal{N}(0, I_n)} \|G\hat{z} - Gz_0\|^2 = \mathbb{E}_{z_0 \sim \mathcal{N}(0, I_n)} \|G\,\mathcal{P}_{G^{\mathrm{T}}A^{\mathrm{T}}}\, z_0 - Gz_0\|^2 = \|G(\mathcal{P}_{G^{\mathrm{T}}A^{\mathrm{T}}} - I_n)\|_F^2$$

where the last equality follows from Lemma 1.

Without loss of generality, it suffices to consider the case when G is diagonal. To see this, let $U\Sigma V^{\mathrm{T}}$ be the SVD of G. Observe that

$$\mathcal{P}_{G^{\mathrm{T}}A^{\mathrm{T}}} = \mathcal{P}_{V\Sigma U^{\mathrm{T}}A^{\mathrm{T}}} = V\,\mathcal{P}_{\Sigma U^{\mathrm{T}}A^{\mathrm{T}}}\,V^{\mathrm{T}} \stackrel{d}{=} V\,\mathcal{P}_{\Sigma A^{\mathrm{T}}}\,V^{\mathrm{T}}$$

where we used Lemma 2 in the second equality and used the rotational invariance of A in the last equality. Thus we have

$$\|G(\mathcal{P}_{G^{\mathrm{T}}A^{\mathrm{T}}} - I_n)\|_F^2 \stackrel{d}{=} \|G(V\,\mathcal{P}_{\Sigma A^{\mathrm{T}}}\,V^{\mathrm{T}} - I_n)\|_F^2.$$

Moreover, note that

$$\|G(V\,\mathcal{P}_{\Sigma A^{\mathrm{T}}}\,V^{\mathrm{T}} - I_n)\|_F^2 = \|\Sigma V^{\mathrm{T}}(V\,\mathcal{P}_{\Sigma A^{\mathrm{T}}}\,V^{\mathrm{T}} - I_n)\|_F^2 = \|\Sigma\,\mathcal{P}_{\Sigma A^{\mathrm{T}}}\,V^{\mathrm{T}} - \Sigma V^{\mathrm{T}}\|_F^2 = \|\Sigma(\mathcal{P}_{\Sigma A^{\mathrm{T}}} - I_n)\|_F^2$$

where we used the unitary invariance of the Frobenius norm in the first and last equality and the orthogonality of V in the second equality. Hence

$$\mathbb{E}_A\|G(\mathcal{P}_{G^{\mathrm{T}}A^{\mathrm{T}}} - I_n)\|_F^2 = \mathbb{E}_A\|\Sigma(\mathcal{P}_{\Sigma A^{\mathrm{T}}} - I_n)\|_F^2 = \mathbb{E}_A\|(I_n - \mathcal{P}_{\Sigma A^{\mathrm{T}}})\Sigma\|_F^2$$

so we assume without loss of generality that $G = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and consider $\mathbb{E}_A\|(I_n - \mathcal{P}_{GA^{\mathrm{T}}})G\|_F^2$.

We now compute the lower bound of the expected error. Note that $\mathrm{rank}(\mathcal{P}_{GA^{\mathrm{T}}}) = m$ since A is full rank with probability 1. Also for any random draw of Gaussian A,

$$\min_{S \in \mathbb{R}^{n \times n},\ \mathrm{rank}(S) \leqslant m} \|G - S\|_F^2 \leqslant \|(I_n - \mathcal{P}_{GA^{\mathrm{T}}})G\|_F^2. \tag{1}$$

By the Eckart-Young Theorem (Eckart & Young, 1936),

$$\min_{S\in\mathbb{R}^{n\times n},\ \mathrm{rank}(S)\leqslant m} \|G-S\|_F^2 = \sum_{i>m}\sigma_i^2.$$

Taking the expectation with respect to A to the right hand-side of (1) establishes the lower bound in the theorem.

The upper bound comes from the following result in (Halko et al., 2011).

Theorem 1 (Minor variant of Theorem 10.5 in (Halko et al., 2011))**.** *Suppose G is a real $\ell\times n$ matrix with singular values $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_{\min\{\ell,n\}}\geqslant 0$. Choose a target rank $k\geqslant 2$ and oversampling factor $m-k\geqslant 2$ where $m\leqslant\min\{\ell,n\}$. Draw an $n\times m$ Gaussian matrix A^{T} and construct the sample matrix GA^{T}. Then the expected approximation error*

$$\mathbb{E}_A\|(I_n-\mathcal{P}_{GA^{\mathrm{T}}})G\|_F^2\leqslant\left(1+\frac{k}{m-k-1}\right)\sum_{j>k}\sigma_j^2. \tag{2}$$

Theorem 10.5 in (Halko et al., 2011) literally states a bound on $\mathbb{E}_A\|(I_n-\mathcal{P}_{GA^{\mathrm{T}}})G\|_F$ but in the proof, the authors show the stronger result (2). The upper bound in Theorem 1 follows by setting $\ell=n$ and $k=m-2$ in Theorem 1 whereby the condition on k requires $4\leqslant m\leqslant n$.

□

The following are two supplementary lemmas used in the proof of Theorem 1. The first shows that the expected value of the ℓ_2 norm of a matrix acting on an isotropic gaussian vector is its frobenius norm.

Lemma 1. *Let $M\in\mathbb{R}^{m\times n}$. Then $\mathbb{E}_{z\sim\mathcal{N}(0,I_n)}\|Mz\|^2=\|M\|_F^2$.*

Proof. Let $U\Sigma V^{\mathrm{T}}$ be the SVD of M where $\Sigma\in\mathbb{R}^{m\times n}$ contains the singular values σ_i of M on the diagonal for $i=1,\ldots,r$ and $r=\mathrm{rank}(M)\leqslant\min\{m,n\}$. Then

$$\|Mz\|^2=\|U\Sigma V^{\mathrm{T}}z\|^2=\|\Sigma V^{\mathrm{T}}z\|^2\overset{d}{=}\|\Sigma z\|^2$$

where we used the unitary invariance of the ℓ_2 norm in the second equality and the rotational invariance of z in the last equality. The result follows by noting that

$$\mathbb{E}_{z\sim\mathcal{N}(0,I_n)}\|\Sigma z\|^2=\sum_{i=1}^r\sigma_i^2\mathbb{E}_{z_i\sim\mathcal{N}(0,1)}z_i^2=\sum_{i=1}^r\sigma_i^2=\|M\|_F^2.$$

□

The second lemma asserts that unitary matrices exhibit a commutativity property when acting on the range of an orthogonal projector.

Lemma 2 (Proposition 8.4 in (Halko et al., 2011))**.** *Given a matrix M, let $\mathcal{P}_M$ denote the orthogonal projection onto the range of M. Then for any unitary U, $U^{\mathrm{T}}\,\mathcal{P}_M\,U=\mathcal{P}_{U^{\mathrm{T}}M}$.*

1.2. Models

1.2.1. Architecture and Training Details

We train 128px and 64px variants of the Glow architecture (Kingma & Dhariwal, 2018). This model uses a sequence of invertible flow steps, each comprising of an activation normalization layer, a 1×1 invertible convolutional layer, and an affine couping layer. Let K be the number of steps of flow before each splitting layer, and let L be the number of times the splitting is performed. For denoising, we use $K = 48$, $L = 4$ for 64px recovery. For compressive sensing, we use $K = 18$ and $L = 4$ for 64px recovery and $K = 32$, $L = 6$ for 128px recovery. All models are trained over 5-bit images using a learning rate of 0.0001 and $10,000$ warmup iterations, as in (Kingma & Dhariwal, 2018). When solving inverse problems using Glow, original 8-bit images were used.

We observed numerical instability when solving inverse problems in the 128px case, in that activations of the Glow network could become too large to compute during inversion. To mitigate this problem, we train a modified version of the Glow reference implementation in which we add a small constant $\epsilon = 0.0005$ to computed scale parameters in each Actnorm layer, thus preventing a "division by zero error" during inversion. Additionally, in forward passes of the Glow model we clip activations to the range $[-40, 40]$.

We train a 64px DCGAN (Radford et al., 2016) and a 128px PGGAN (Karras et al., 2018) in order to compare the Glow model to traditional GAN architectures. The DCGAN model has $d = 5$ upsampling layers implemented through transpose convolutions. The PGGAN model has $d = 5$ upsampling layers implemented through nearest neighbor upsampling followed by a convolutional layer. In both cases, the first layer of each GAN includes $k_0 = 512$ activation channels and after each upsampling operation the number of activation channels is reduced by half. These architectures are used as-is for both the CSGM (Bora et al., 2017) and the Image Adaptive (Hussein et al., 2020) compressive sensing procedures.

Lastly, we also use an overparametrized variant of the Deep Decoder (Heckel & Hand, 2019) in order to compare to an unlearned neural network for compressive sensing. We use $d = 5$ and $d = 6$ upsampling layers for the 64px and 128px cases respectively, implemented in both cases with convolutional layers followed by bilinear upsampling. The number of activation channels in each layer is held fixed at $k = 250$ and $k = 700$ for the 64px and 128px cases respectively.

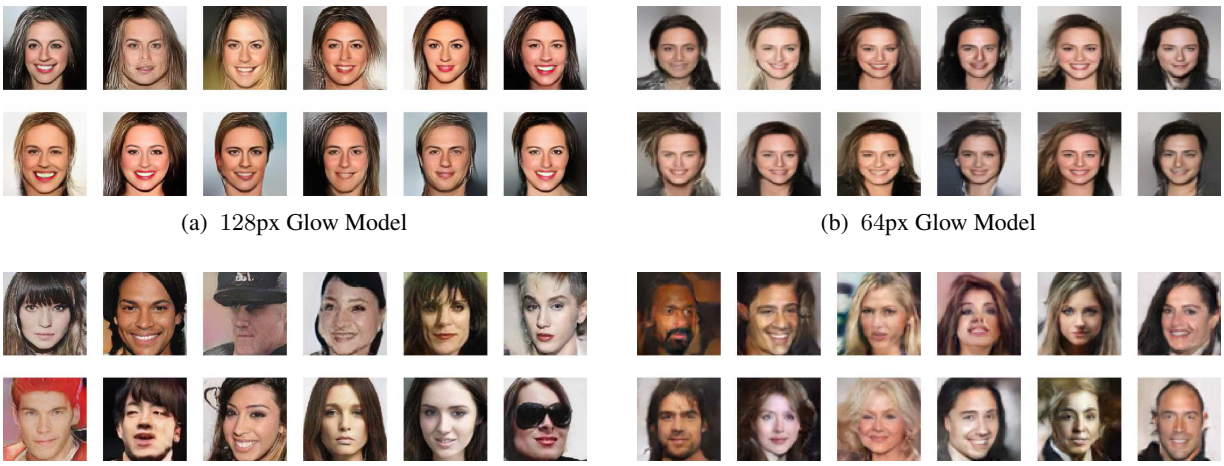
64px Models	Hyperparameters	Repr. Size	Model Size	Overparam. Ratio
Glow (Denoising)	$K = 48$, $L = 4$, affine coupling	12,288	67,172,224	1.000
Glow (CS)	$K = 18$, $L = 4$, affine coupling	12,288	25,783,744	1.000
DCGAN	$k_0 = 512$, $d = 5$	100	3,576,804	0.0081
IA-DCGAN	$k_0 = 512$, $d = 5$	3,576,804	3,576,804	291.1
Deep Decoder	$k = 250$, $d = 5$	253,753	254,753	20.65

128px Models	Hyperparameters	Repr. Size	Model Size	Overparam. Ratio
Glow	$K = 32$, $L = 6$, affine coupling	49152	129,451,520	1.000
PGGAN	$k_0 = 512$, $d = 5$	512	13,625,904	0.0104
IA-PGGAN	$k_0 = 512$, $d = 5$	13,625,904	13,625,904	277.2
Deep Decoder	$k = 700$, $d = 6$	2,462,603	2,473,803	50.10

Figure 1. Summary of the model parameters for all models used in our experiments. Hyperparameters refer to model-specific configurations as described in the text. Representation Size (Repr. Size) is the dimensionality of each model's image representation, i.e. the total number of optimizable parameters available during inversion in each inverse problem. The model size is the total sum of all parameters of each model, including those of the image representation. The Overparametrization Ratio (Overparam. Ratio) describes the representation size as a fraction of the output image dimensionality.

1.2.2. Unconditional Output Samples

To demonstrate the successful training of our generative models, we provide output samples from each of the learned models used in our experiments.



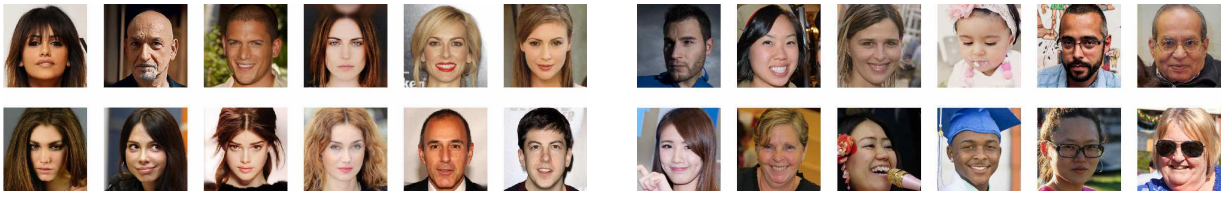
(a) 128px Glow Model (b) 64px Glow Model

(c) PGGAN Model (d) DCGAN Model

Figure 2. Samples from each of our trained generative models.

1.2.3. Test Set Images

Throughout our paper, we present experiments across two test sets, one from the same distribution of the trained generative models and another which is out-of-distribution. In particular, our in-distribution test set is sampled randomly from a validation split of the CelebA-HQ dataset (Karras et al., 2018) and our out-of-distribution test set is sampled randomly from the Flickr Faces High Quality (FFHQ) Dataset (Karras et al., 2019). None of the test set images are seen by the generative models during their training phases. A few samples from both test sets are shown below.



(a) CelebA-HQ Test Set (b) FFHQ Test Set

Figure 3. A few samples from each of the test sets used in our experiments.

1.3. Denoising: Additional Experiments

We additionally provide qualitative and quantitative denoising results over the test set of in-distribution CelebA images in Figures 4 and 5. Recall that we formulate denoising as the following empirical risk minimization problem:

$$\min_{z \in \mathbb{R}^n} \|G(z) - y\|^2 + \gamma \|z\|^2 \qquad y = x_0 + \eta$$

Where $\eta \sim \mathcal{N}(0, \sigma^2 I_n)$ is additive Gaussian noise. PSNRs for varying choices of the penalization parameter γ under noise levels $\sigma = 0.01, 0.05, 0.1$, and 0.2 are presented in Figure 4 below.

The central message is that the Glow prior outperforms the DCGAN prior uniformly across all γ due to the representation error of the DCGAN prior. In addition, when γ is chosen appropriately, regularization improves the performance of Glow, which can outperform the state-of-the-art BM3D algorithm at high noise levels such as $\sigma = 0.2$, and can offer comparable performance at lower noise levels. This is in contrast to the DCGAN prior, whose performance is harmed by increased regularization.

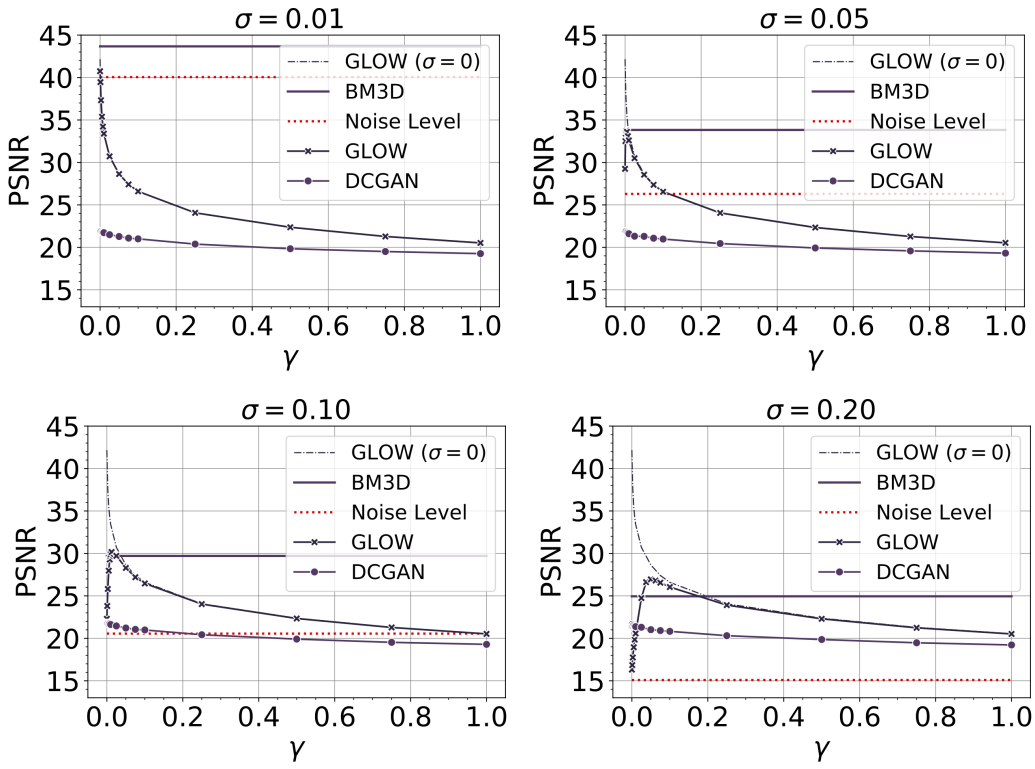


Figure 4. Image Denoising — Recovered PSNR values as a function of γ on $N = 50$ in-distribution test set CelebA images. We report the average PSNR after applying BM3D, under the DCGAN prior, and under the Glow prior, at noise levels $\sigma = 0.01, 0.05, 0.10, 0.20$. For reference, we also show the average PSNR of the original noisy images and the average PSNR of images recovered by the Glow prior in the noiseless case ($\sigma = 0$).

Figure 5. Image Denoising — Visual comparisons under the Glow prior, the DCGAN prior, and BM3D at noise level $\sigma = 0.2$ on a sample of CelebA in-distribution test set images. Under the DCGAN prior, we only show the case of $\gamma = 0$ as this consistently gives the best performance. Under the Glow prior, the best performance is achieved with $\gamma = 0.05$, overfitting of the image occurs with $\gamma = 0.001$ and underfitting occurs with $\gamma = 1$. Note that the Glow prior with $\gamma = 0.05$ also gives a sharper image than BM3D.

1.4. Compressive Sensing: Additional Experiments

We first analyze our formulation of compressive sensing as an empirical risk minimization problem. Recall that we solve the following optimization problem:

$$\min_{z \in \mathbb{R}^n} \|AG(z) - y\|^2 \qquad y = Ax_0 + \eta$$

Where A is a Gaussian random measurement matrix with i.i.d. $\mathcal{N}(0, 1/m)$ entries, and η is noise normalized so that $\sqrt{\mathbb{E}\|\eta\|^2} = 0.1$. This formulation includes no explicit penalization on the likelihood of the latent code z, and instead relies on implicit regularization through the use of a gradient based optimization method initialized at $z_0 = 0$. To justify this formulation, we study various alternative methods of initialization: $z_0 = 0$, $z_0 \sim \mathcal{N}(0, 0.1^2 I_n)$, $z_0 \sim \mathcal{N}(0, 0.7^2 I_n)$, $z_0 = G^{-1}(x_0)$ with x_0 given by the solution to Lasso with respect to the wavelet basis, and $z_0 = G^{-1}(x_0)$ where x_0 is perturbed by a random point in the null space of A. For each initialization, we plot recovery performance on a compressive sensing task as a function of γ for the regularized objective:

$$\min_{z \in \mathbb{R}^n} \|AG(z) - y\|^2 + \gamma \|z\|^2$$

As shown in Fig. 6, the Glow model shows best performance with $\gamma = 0$ with initialization at the zero vector, despite there being no explicit penalization on the likelihood of recovered latent codes.

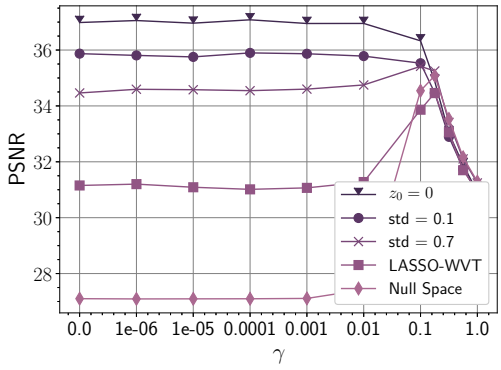


Figure 6. Performance of the Glow model in Compressive Sensing under various initialization strategies (as described in the text). The Glow model shows strongest performance when initialized with the zero vector and with no explicit penalization on the latent code likelihood. The task is 64px CelebA image recovery using $m = 5000$ ($\approx 50\%$) Gaussian measurements.

We show in Fig. 7 the results of our compressive sensing recovery experiments using the Structural Similarity index (SSIM) as a recovery quality metric (Wang et al., 2004). SSIM is designed to indicate perceptual quality by extracting structural information from images.

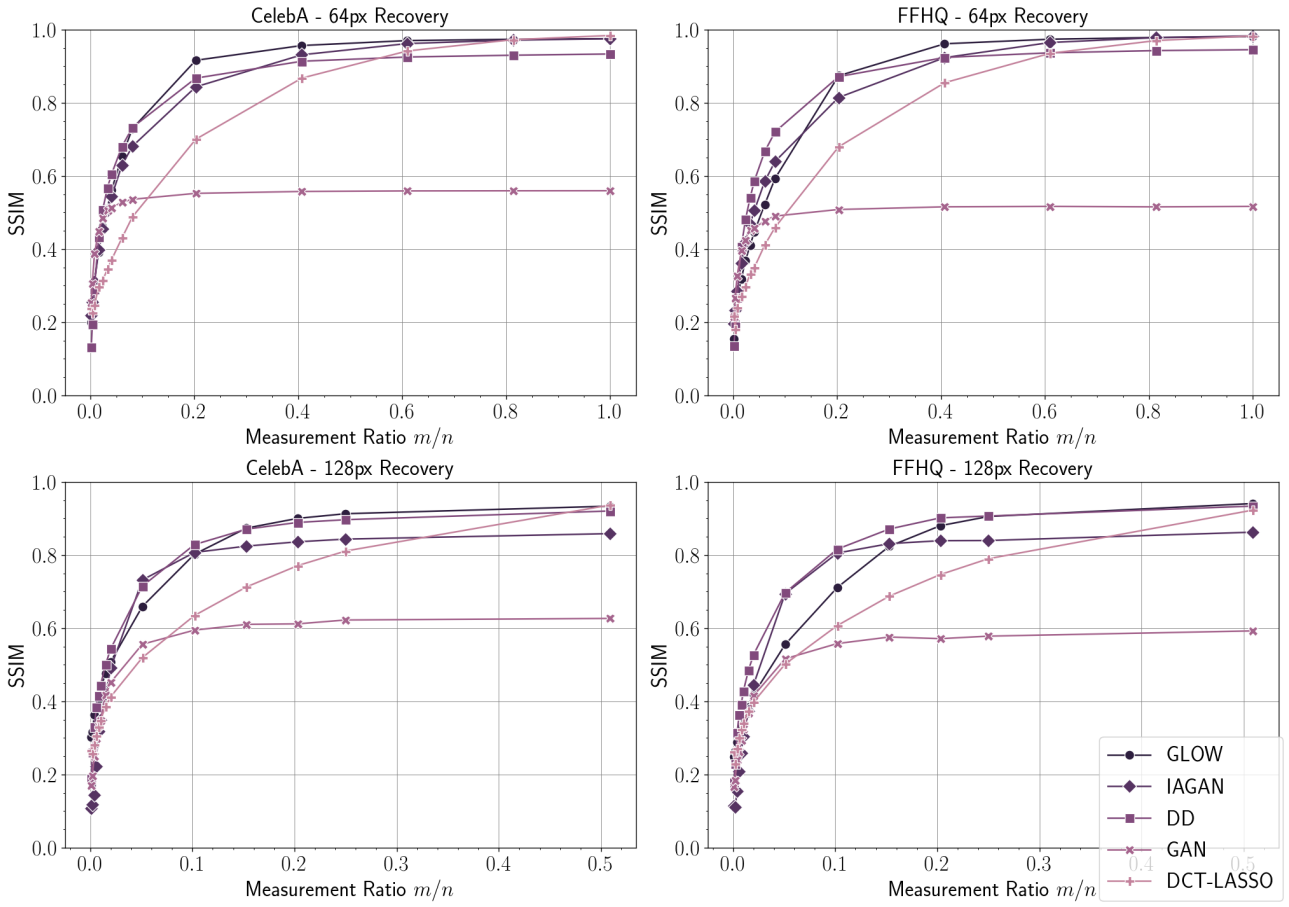


Figure 7. 64px and 128px Compressive Sensing results using $N = 1000$ and $N = 100$ test set images respectively. Each plot shows SSIM (higher is better, ranging in $[0, 1]$) across a variety of measurement ratios.

1.4.1. Compressive Sensing 64px Sample Sheets

We provide here additional sample sheets for our 64px experiments in compressive sensing. For in-distribution images, we show in Figures 8 and 9 qualitative examples of the image reconstructions for the Glow prior, the DCGAN prior, the IA-DCGAN prior, and the Deep Decoder. We replicate the same experiments for out-of-distribution FFHQ images in Figures 10 and 11.

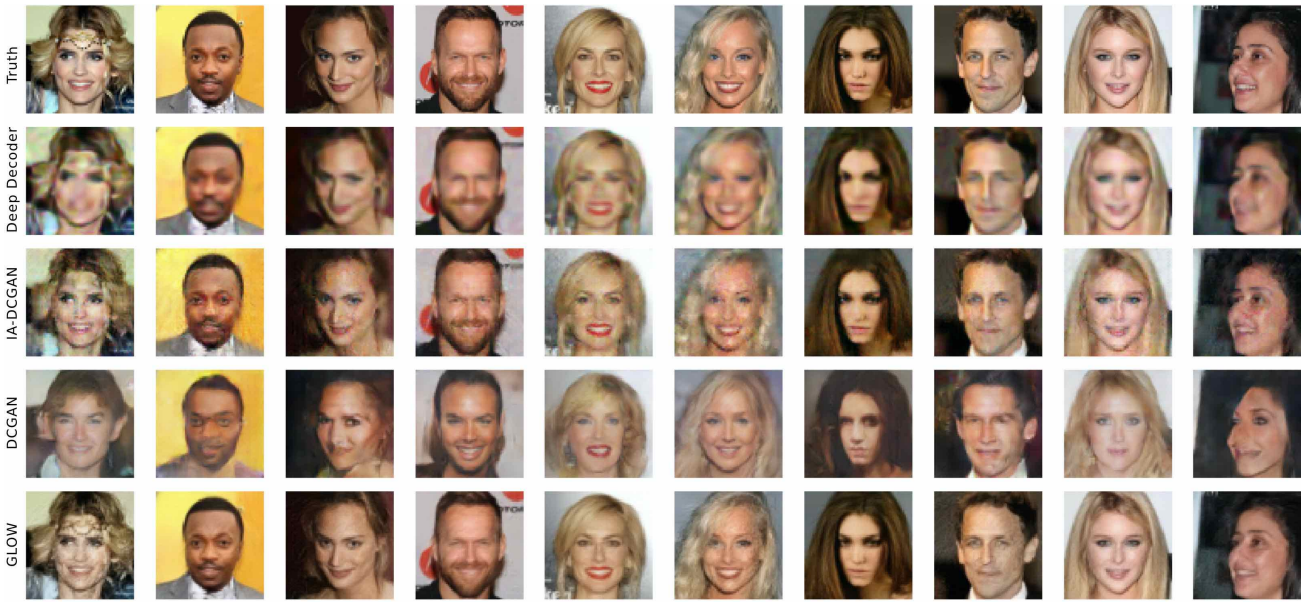


Figure 8. Compressive sensing visual comparisons — Recoveries on a sample of in-distribution test set images with a number $m = 2500$ ($\approx 20\%$) of measurements under the Glow prior, the DCGAN prior, the IA-DCGAN prior, and the Deep Decoder.

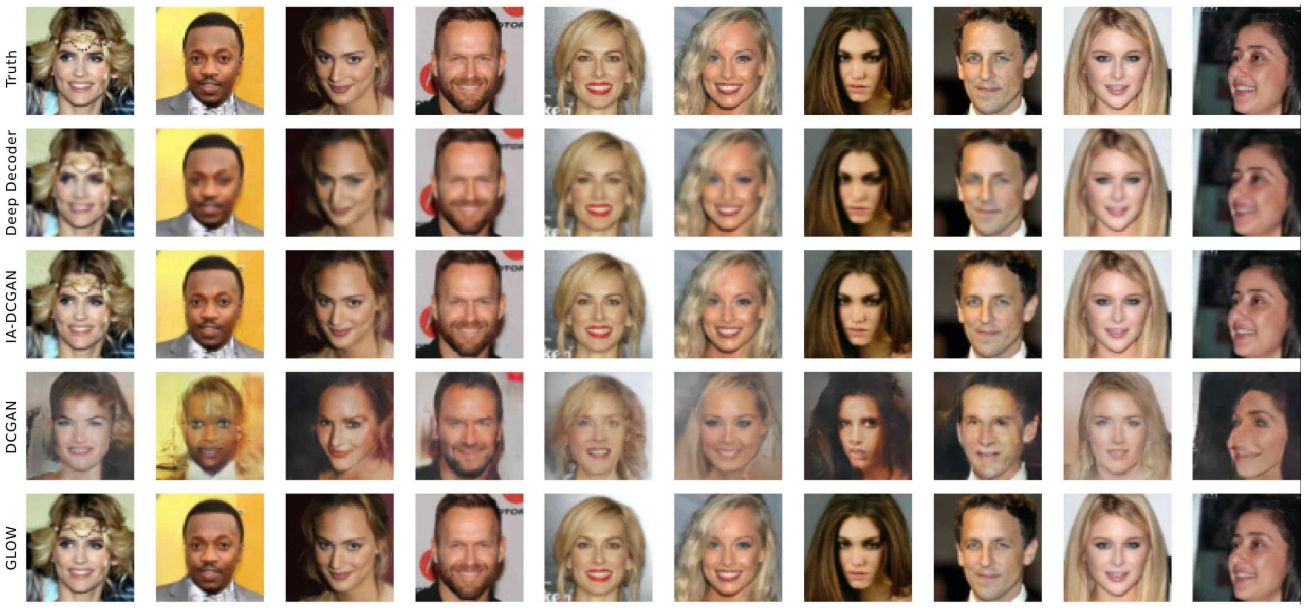


Figure 9. Compressive sensing visual comparisons — Recoveries on a sample of in-distribution test set images with a number $m = 10000$ ($\approx 80\%$) of measurements under the Glow prior, the DCGAN prior, the IA-DCGAN prior, and the Deep Decoder.

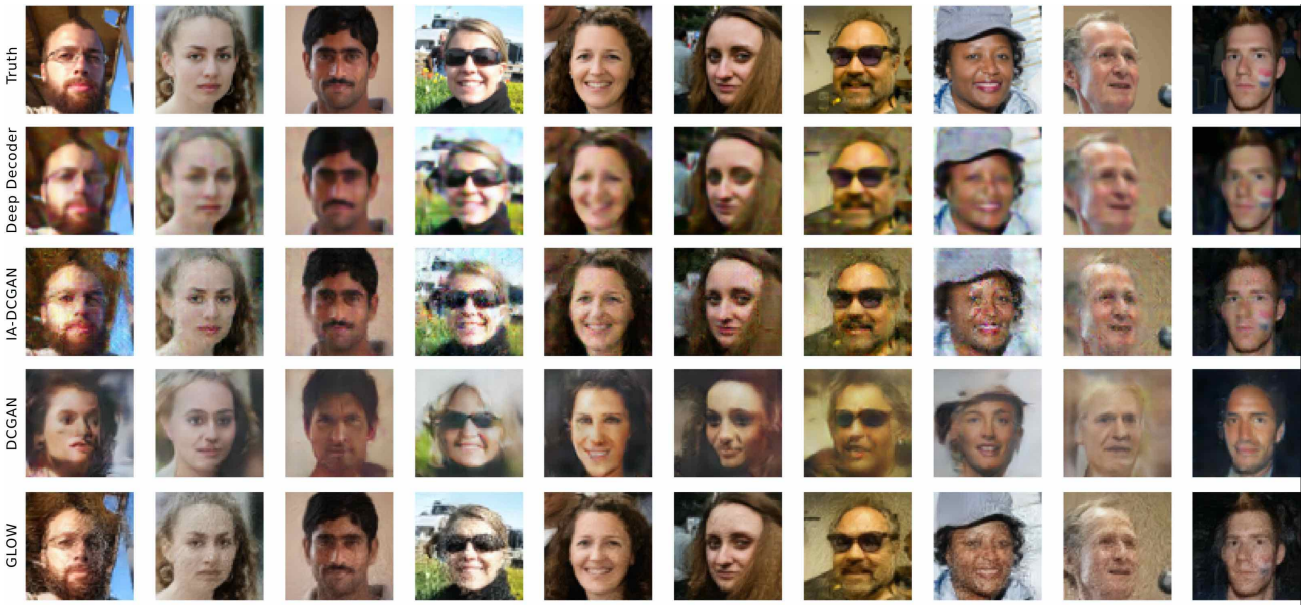


Figure 10. Compressive sensing visual comparisons — Recoveries on a sample of out-of-distribution (FFHQ) test set images with a number $m = 2500$ ($\approx 20\%$) of measurements under the Glow prior, the DCGAN prior, the IA-DCGAN prior, and the Deep Decoder.

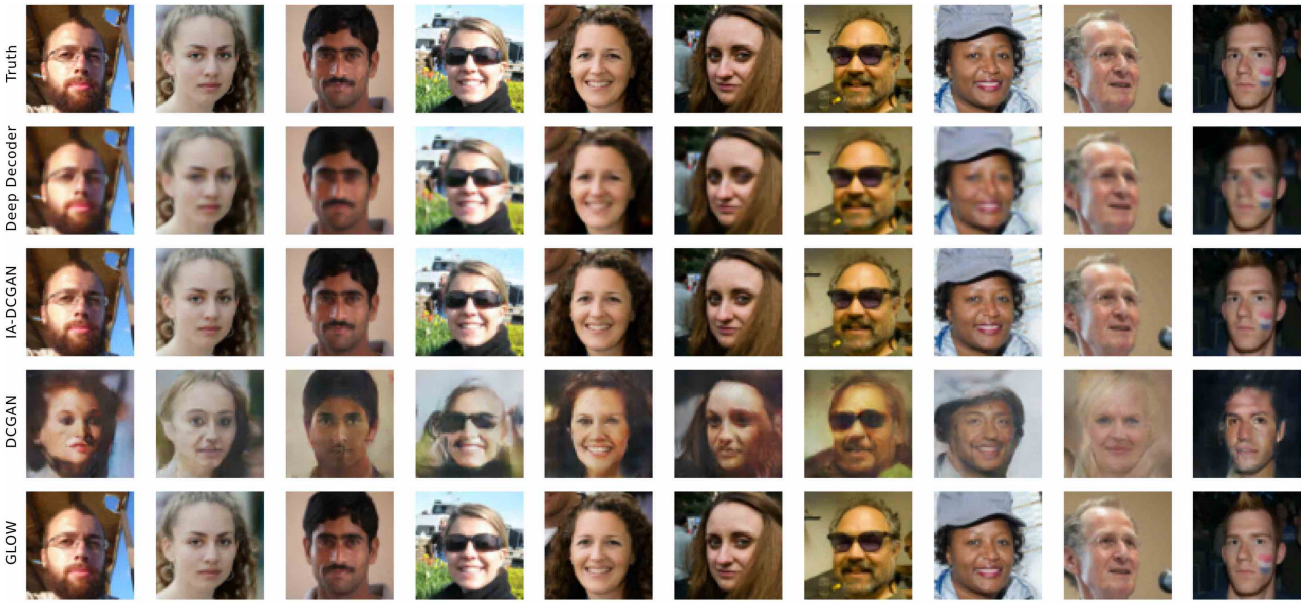


Figure 11. Compressive sensing visual comparisons — Recoveries on a sample of out-of-distribution (FFHQ) test set images with a number $m = 10000$ ($\approx 80\%$) of measurements under the Glow prior, the DCGAN prior, the IA-DCGAN prior, and the Deep Decoder.

To further investigate whether the Glow prior continues to be an effective proxy for arbitrarily out-of-distribution images, we tested arbitrary natural images such as a car, house door, and butterfly wings, which are all semantically unrelated to CelebA images. In general, we found that Glow is an effective prior at compressive sensing of out-of-distribution natural images, which are assigned a relatively high likelihood score (small normed latent representations) as compared to noisy, unnatural images. On these images, Glow also outperforms LASSO. This means that invertible networks have at least partially learned something more general about natural images from the CelebA dataset – there may be some high level features that face images share with other natural images, such as smooth regions followed by discontinuities, etc. This allows the Glow model to extend its effectiveness as a prior to other natural images beyond just the training set.

As compared to in-distribution training images, however, semantically unrelated images are assigned very low-likelihood scores by the Glow model, causing instability issues. In particular, an L-BFGS search for the solution of an inverse problem to recover a low-likelihood image leads the iterates into neighborhoods of low-likelihood representations that may induce instability. All the network parameters such as scaling in the coupling layers of Glow network are learned to behave stably with high likelihood representations. However, on very low-likelihood representations, unseen during the training process, the networks becomes unstable and outputs of network begin to diverge to very large values; this may be due to several reasons, such as normalization (scaling) layers not being tuned to the unseen representations. See Section 1.2.1 for details on our approach to handing these instabilities.

We show in Figure 12 a comparison of the performance of the LASSO-DCT, LASSO-WVT, DCGAN, and Glow priors on the compressive sensing of 64px out-of-distribution images for $m = 2500$ ($\approx 20\%$) measurements.

Figure 12. Compressive sensing of 64px out-of-distribution images given $m = 2500$ ($\approx 20\%$) measurements at a noise level $\sqrt{\mathbb{E}\|\eta\|^2} = 0.1$. We provide a visual comparison of recoveries of the LASSO-DCT, LASSO-WVT, DCGAN, and Glow priors, where the DCGAN and Glow priors are trained on CelebA images. In each case, we choose values of the penalization parameter γ to yield the best performance. We use $\gamma = 0$ for both the DCGAN and Glow priors, and optimize γ for each recovery using the LASSO-WVT and LASSO-DCT priors.

1.4.2. Compressive Sensing 128px Sample Sheets

We provide here additional sample sheets for our 128px experiments in compressive sensing. For in-distribution images, we show in Figures 13 and 14 qualitative examples of the image reconstructions for the Glow prior, the PGGAN prior, the IA-PGGAN prior, and the Deep Decoder. We replicate the same experiments for out-of-distribution (FFHQ) images in Figures 15 and 16.

Figure 13. Compressive sensing visual comparisons — Recoveries on a sample of in-distribution test set images with a number $m = 10000$ ($\approx 20\%$) of measurements under the Glow prior, the PGGAN prior, the IA-PGGAN prior, and the Deep Decoder.

Figure 14. Compressive sensing visual comparisons — Recoveries on a sample of in-distribution test set images with a number $m = 2500$ ($\approx 5\%$) of measurements under the Glow prior, the PGGAN prior, the IA-PGGAN prior, and the Deep Decoder.

Figure 15. Compressive sensing visual comparisons — Recoveries on a sample of out-of-distribution (FFHQ) test set images with a number $m = 10000$ ($\approx 20\%$) of measurements under the Glow prior, the PGGAN prior, the IA-PGGAN prior, and the Deep Decoder.

Figure 16. Compressive sensing visual comparisons — Recoveries on a sample of out-of-distribution (FFHQ) test set images with a number $m = 2500$ ($\approx 5\%$) of measurements under the Glow prior, the PGGAN prior, the IA-PGGAN prior, and the Deep Decoder.

1.5. Inpainting

We present here qualitative results on image inpainting under the DCGAN prior and the Glow prior on the CelebA test set. Compared to DCGAN, the reconstructions from Glow are of noticeably higher visual quality.

Figure 17. Image inpainiting results on a sample of the CelebA test set. Masked images are recovered using the DCGAN prior and the Glow prior. Recoveries under the DCGAN prior are skewed and blurred whereas the Glow prior leads to sharper and coherent inpainted images. For both Glow and DCGAN, we set $\gamma = 0$.

1.5.1. Image Inpainting on Out of Distribution Images

We now perform image inpainiting under the Glow prior and the DCGAN prior, each trained on CelebA. Figure 18 shows the visuals of out-of-distribution inpainting. As before, the DCGAN prior continues to suffer due to representation error and data bias while Glow achieves reasonable reconstructions on out-of-distribution images which are semantically similar to CelebA faces. As one deviates to other natural images such as houses, doors, and butterfly wings, the inpainting performance deteriorates. In compressive sensing, Glow performs much better on such arbitrarily out-of-distribution images as in this case, good recoveries only require the network only to assign a higher likelihood score to the true image compared to the all the candidate static images given by the null space of the measurement operator.

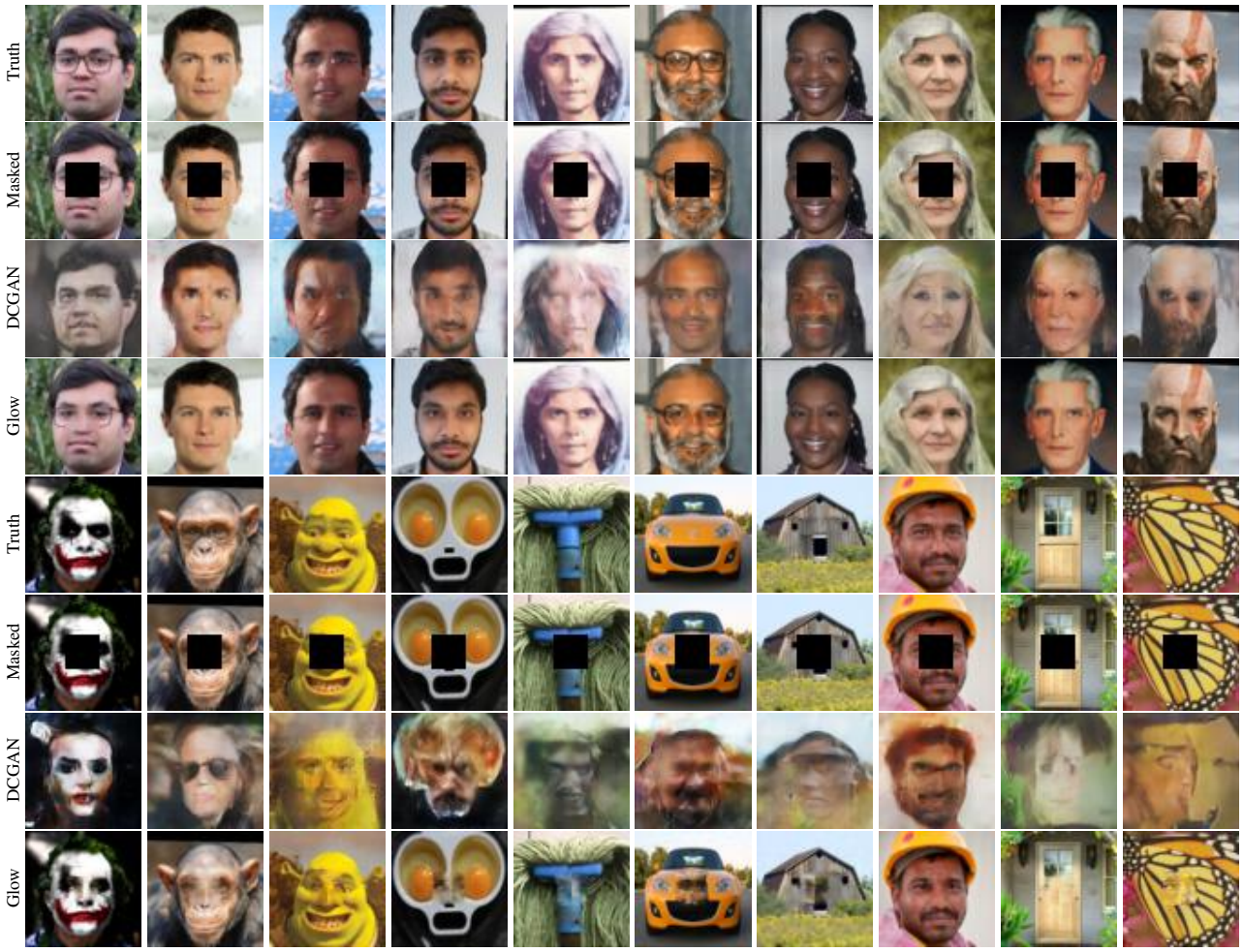


Figure 18. Image inpainiting results on out-of-distribution images. Masked images are recovered under the DCGAN prior and the Glow prior. Recoveries under the DCGAN prior are skewed and blurred whereas Glow prior leads to sharper and coherent inpainted images. For both Glow and DCGAN, we set $\gamma = 0$.

1.6. Discussion

We provide evidence that random perturbations in image space induce larger changes in z than comparable natural perturbations in image space. Figure 19 shows a plot of the norm of the change in image space, averaged over 100 test images, as a function of the size of a perturbation in latent space. Natural directions are given by the interpolation between the latent representation of two test images. For the denoising problem, this difference in sensitivity indicates that the optimization algorithm might obtain a larger decrease in $\|z\|$ by an image modification that reduces unnatural image components than by a correspondingly large modification in a natural direction.

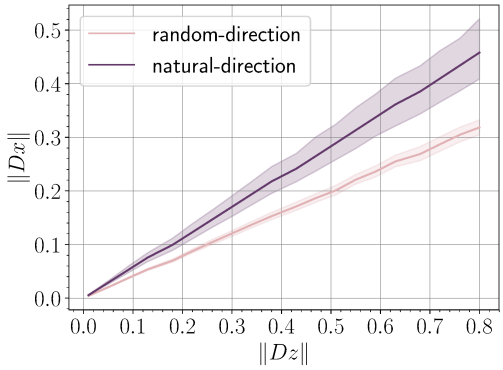


Figure 19. The magnitude of the change in image space as a function of the size of a perturbation in latent space. Solid lines are the mean behavior and shaded region depicts 95% confidence interval.

To further illustrate this point, we investigate in Figure 20 the decay of the singular values of the Glow model's Jacobian for random points.

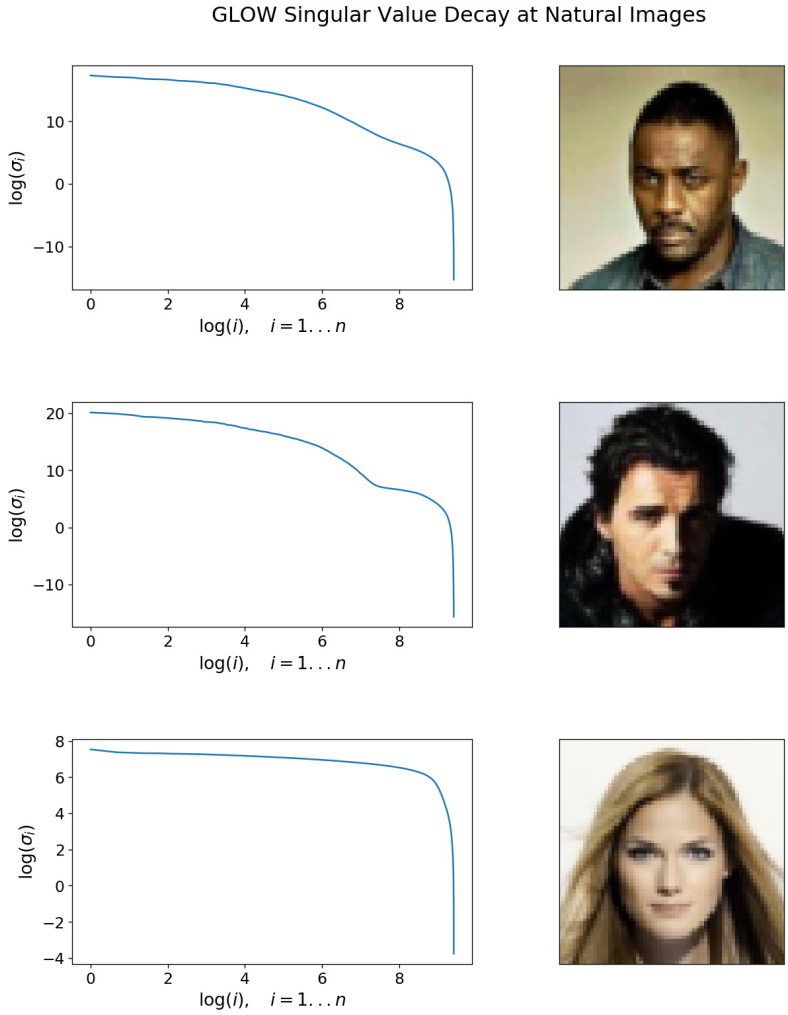


Figure 20. Log-log plots of the decay of the singular values of the trained Glow model's Jacobian for three random CelebA images.

References

Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 537–546. PMLR, Aug 2017.

Eckart, C. and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

Halko, N., Martinsson, P.-G., and Tropp, J. A. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53:217–288, 2011.

Heckel, R. and Hand, P. Deep decoder: Concise image representations from untrained non-convolutional networks. In *International Conference on Learning Representations*, 2019.

Hussein, S. A., Tirer, T., and Giryes, R. Image-adaptive gan based reconstruction. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 3121–3129. AAAI Press, 2020.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 4401–4410. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.00453.

Kingma, D. P. and Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. In Bengio, S., Wallach, H. M., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada*, pp. 10236–10245, 2018.

Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Bengio, Y. and LeCun, Y. (eds.), *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, Apr 2004. ISSN 1941-0042. doi: 10.1109/TIP.2003.819861.