
Optimizing the Latent Space of Generative Networks

Piotr Bojanowski, Armand Joulin, David Lopez-Paz, Arthur Szlam

Facebook AI Research

{bojanowski, ajoulin, dlp, aszlam}@fb.com

Abstract

Generative Adversarial Networks (GANs) have been shown to be able to sample impressively realistic images. GAN training consists of a saddle point optimization problem that can be thought of as an adversarial game between a generator which produces the images, and a discriminator, which judges if the images are real. Both the generator and the discriminator are commonly parametrized as deep convolutional neural networks. The goal of this paper is to disentangle the contribution of the optimization procedure and the network parametrization to the success of GANs. To this end we introduce and study *Generative Latent Optimization* (GLO), a framework to train a generator without the need to learn a discriminator, thus avoiding challenging adversarial optimization problems. We show experimentally that GLO enjoys many of the desirable properties of GANs: learning from large data, synthesizing visually-appealing samples, interpolating meaningfully between samples, and performing linear arithmetic with noise vectors.

1 Introduction

Generative Adversarial Networks (GANs) [Goodfellow et al., 2014] are a powerful framework to learn generative models of natural images. GANs learn these generative models by setting up an adversarial game between two learning machines. First, a *generator* plays to transform *noise vectors* into *fake samples*, which resemble *real samples* drawn from a distribution of natural images. Second, a *discriminator* plays to distinguish between real and fake samples. During training, the generator and the discriminator learn in turns. First, the discriminator learns to assign high scores to real samples, and low scores to fake samples. Second, the generator learns to increase the scores of fake samples, as to fool the discriminator. After proper training, the generator is able to produce visually-appealing samples from noise vectors.

Recently, GANs have been used to produce samples of remarkable quality from distributions of natural images, such as handwritten digits, human faces, and house interiors [Radford et al., 2015]. Furthermore, GANs exhibit three strong signs of generalization. First, the generator translates *linear interpolations in the noise space* into *semantic interpolations in the image space*. In other words, a linear interpolation in the noise space will generate a smooth interpolation of visually-appealing images. Second, the generator allows *linear arithmetic in the noise space*. Similarly to word embeddings [Mikolov et al., 2013], linear arithmetic indicates that the generator organizes the noise space to disentangle the nonlinear factors of variation of natural images into linear statistics. Third, the generator is able to to synthesize new visually-appealing samples. This allows for applications such as image inpainting [Iizuka et al., 2017] and super-resolution [Ledig et al., 2016].

Despite their successes, training and evaluating GANs is notoriously difficult. The adversarial optimization problem implemented by GANs is sensitive to random initialization, architectural choices, and hyper-parameter settings. In many cases, a fair amount of human care is necessary to find the correct configuration to train a GAN in a particular dataset. It is common to observe generators with similar architectures and hyper-parameters to exhibit dramatically different behaviors. Even when properly trained, the resulting generator may synthesize samples that resemble only a few

localized regions (or modes) of the data distribution [Goodfellow, 2017]. While several advances have been made to stabilize the training of GANs [Salimans et al., 2016], this task remains more art than science.

The difficulty of training GANs is aggravated by the challenges in their evaluation: since the likelihood of a GAN with respect to the data is intractable, the current gold standard to evaluate the quality of a GAN model is to eyeball the samples produced by the generator. The evaluation of the discriminator is also not simple, since the visual features learned by the discriminator do not always transfer well to supervised tasks [Donahue et al., 2016, Dumoulin et al., 2016]. Finally, the application of GANs to non-image data has been relatively limited.

Research question To model natural images with GANs, the generator and the discriminator are often parametrized as deep Convolutional Networks (convnets) [LeCun et al., 1998a]. Therefore, it is reasonable to hypothesize that the reasons for the success of GANs in modeling natural images come from two complementary sources:

- (A1) Leveraging the powerful inductive bias of deep convnets.
- (A2) The adversarial training protocol.

This work attempts to disentangle the factors of success (A1) and (A2) in GAN models. Specifically, we build an algorithm that relies on (A1), avoids (A2), and obtains competitive results when compared to a GAN.

Contribution. We investigate the importance of (A1) and propose a drastic simplification of GAN training (Section 2). Our approach, called *Generative Latent Optimization* (GLO), maps one learnable noise vector to each of the images in our dataset by minimizing a simple reconstruction loss. Since we are predicting *images from learnable noise*, GLO borrows inspiration from recent methods to predict *learnable noise from images* [Bojanowski and Joulin, 2017]. Alternatively, one can understand GLO as an auto-encoder where the latent representation is not produced by a parametric encoder, but learned freely. In contrast to GANs, we track of the correspondence between each learned noise vector and the image that it represents. Hence, the goal of GLO is to find a meaningful organization of the noise vectors, such that they can be mapped to their target images. To turn GLO into a generative model, we observe that it suffices to draw vectors from the span of the learned noise vectors.

In our experiments (Section 3), we show that GLO inherits many of the appealing features of GANs, while enjoying much simpler training. In particular, we study the efficacy of GLO to compress and decompress a dataset of images (Section 3.3.1), generate new samples (Section 3.3.2), perform linear interpolations and extrapolations in the noise space (Section 3.3.3), perform linear arithmetics (Section 3.3.4) and super-resolution (Section 3.3.5). Our experiments provide quantitative and qualitative comparisons to Principal Component Analysis (PCA), Deep Convolutional Auto-Encoders (DCAE), and GANs. We focus on the MNIST, CIFAR-10, STL-10, CelebA, and LSUN-Bedroom datasets. We conclude our exposition in Section 4.

2 The Generative Latent Optimization (GLO) model

First, we consider a large set of images $\{x_1, \dots, x_N\}$, where each image $x_i \in \mathcal{X}$ has dimensions $3 \times w \times h$. Second, we initialize a set of d -dimensional random vectors $\{z_1, \dots, z_N\}$, where $z_i \in \mathcal{Z} \subseteq \mathbb{R}^d$ for all $i = 1, \dots, N$. Third, we pair the dataset of images with the random vectors, obtaining the dataset $\{(z_1, x_1), \dots, (z_N, x_N)\}$. Finally, we jointly learn the parameters θ in Θ of a generator $g_\theta : \mathcal{Z} \rightarrow \mathcal{X}$ and the optimal noise vector z_i for each image x_i , by solving:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N \left[\min_{z_i \in \mathcal{Z}} \ell(g_\theta(z_i), x_i) \right], \quad (1)$$

In the previous, $\ell : \mathcal{X} \times \mathcal{X}$ is a loss function measuring the reconstruction error from $g(z_i)$ to x_i . We call this model Generative Latent Optimization (GLO). Next, let us describe the most distinctive features of GLO.

Learnable z_i . One of the distinctive features of GLO is the joint optimization of the random inputs z_1, \dots, z_N and the model parameter θ . This is in contrast to autoencoders [Bourlard and Kamp, 1988], which assume a parametric model $f : \mathcal{X} \rightarrow \mathcal{Z}$, usually referred to as *encoder*, to compute the vector z from samples x , and minimize the reconstruction loss $\ell(g(f(x)), x)$. Since in GLO the vector z is a free parameter, our model can recover all the solutions to be found by an autoencoder, and reach some others. In a nutshell, GLO can be viewed as an “encoder-less” autoencoder, or as a “discriminator-less” GAN.

Choice of \mathcal{Z} . The representation space \mathcal{Z} should encapsulate all of our *prior knowledge* about the data $\{x_1, \dots, x_N\}$. For instance, in the sparse coding literature, it is common practice to set the representation space to the closed unit ball $\mathcal{B}(1, d, p)$ [Bach et al., 2011], where

$$\mathcal{Z} = \mathcal{B}(r, d, p) = \{z \in \mathbb{R}^d : \|z\|_p \leq r\}.$$

Since we are interested in matching the the properties of GANs, we make similar choices to them when it comes to the representation space \mathcal{Z} . The most common choices of the representation space for GANs are either the uniform distribution on the hypercube $[-1, +1]^d$, or the Normal distribution on \mathbb{R}^d . Since Gaussian distributions lead to more stable GAN training [Radford et al., 2015], we will use this distribution to design our representation space. Since random vectors z drawn from the d -dimensional Normal distribution are very unlikely to land far outside the ball $\mathcal{B}(\sqrt{d}, d, 2)$, we rescale random vectors drawn from the d -dimensional Normal distribution, and set our representation space to $\mathcal{B}(1, d, 2)$.

Choice of loss function. On the one hand, the squared-loss function $\ell_2(x, x') = \|x - x'\|_2^2$ is a simple choice, but leads to blurry reconstructions of natural images. On the other hand, GANs use a convnet (the discriminator) as loss function. Since the early layers of convnets focus on edges, the samples from a GAN are sharper. Therefore, our experiments provide quantitative and qualitative comparisons between the ℓ_2 loss and the Laplacian pyramid Lap_1 loss [Ling and Okada, 2006]. The Lap_1 loss corresponds to a small convnet that estimates the Earth mover’s distance in the pixel space, using an Euclidean ground metric. Mathematically,

$$\text{Lap}_1(x, x') = \sum_j 2^{-2j} |L^j(x) - L^j(x')|_1$$

where $L^j(x)$ is the j -th level of the Laplacian pyramid representation of x . Therefore, the Lap_1 loss weights the details at multiple scales equally, and in particular, does not penalize small shifts in the locations of edges as strongly as the Euclidean metric.

Optimization. For any choice of differentiable generator, the objective (1) is differentiable with respect to z , and θ . Therefore, we will learn z and θ by Stochastic Gradient Descent (SGD). The gradient of (1) with respect to z can be obtained by backpropagating the gradients through the generator function [Bora et al., 2017]. We project each z back to the representation space \mathcal{Z} after each update. In the case of $\mathcal{Z} = \mathcal{B}_d^2$, projecting z after each update is a division by $\max(\|z\|_2, 1)$.

2.1 Prior work

Generative Adversarial Networks. GANs were introduced by Goodfellow et al. [2014], and refined in multiple recent works [Denton et al., 2015, Radford et al., 2015, Zhao et al., 2016, Salimans et al., 2016]. As described in Section 1, GANs construct a generative model of a probability distribution P by setting up an adversarial game between a generator g and a discriminator d :

$$\min_G \max_D \mathbb{E}_{x \sim P} \log d(x) + \mathbb{E}_{z \sim Q} (1 - \log d(g(z))).$$

In practice, most of the applications of GANs concern modeling distributions of natural images. In these cases, both the generator g and the discriminator d are parametrized as deep convnets [LeCun et al., 1998a]. Among the multiple architectural variations explored in the literature, the most prominent is the Deep Convolutional Generative Adversarial Network (DCGAN) [Radford et al., 2015]. Therefore, in this paper we will use the specification of the generator function of the DCGAN to construct the generator of GLO across all of our experiments.

Autoencoders. In their simplest form, an Auto-Encoder (AE) is a pair of neural networks, formed by an encoder $f : \mathcal{X} \rightarrow \mathcal{Z}$ and a decoder $g : \mathcal{Z} \rightarrow \mathcal{X}$. The role of an autoencoder is to compress the data $\{x_1, \dots, x_N\}$ into the representation $\{z_1, \dots, z_N\}$ using the encoder $f(x_i)$, and decompress it using the decoder $g(f(x_i))$. Therefore, autoencoders minimize $\mathbb{E}_{x \sim P} \ell(g(f(x)), x)$, where $\ell : \mathcal{X} \times \mathcal{X}$ is a simple loss function, such as the mean squared error. There is a vast literature on autoencoders, spanning three decades from their conception [Bourlard and Kamp, 1988, Baldi and Hornik, 1989], renaissance [Hinton and Salakhutdinov, 2006], and recent probabilistic extensions [Vincent et al., 2008, Kingma and Welling, 2013]. One of the main messages of this paper is that Deep Convolutional Auto-Encoders (DCAE), that is, autoencoders where both the encoder and decoder are deep convnets, exhibit many of the attractive features of GANs but, like GLO, allow a much simpler training.

Several works have combined GANs with AEs. For instance, Zhao et al. [2016] replace the discriminator of a GAN by an AE, and Ulyanov et al. [2017] replace the decoder of an AE by a generator of a GAN. Similar to GLO, these works suggest that the combination of standard pipelines can lead to good generative models. In this work we attempt one step further, to explore if learning a generator alone is possible.

Inverting generators. Several works attempt at recovering the latent representation of an image with respect to a generator. In particular, Lipton and Tripathi [2017], Zhu et al. [2016] show that it is possible to recover z from a generated sample. Similarly, Creswell and Bharath [2016] show that it is possible to learn the inverse transformation of a generator. These works are similar to [Zeiler and Fergus, 2014], where the gradients of a particular feature of a convnet are back-propagated to the pixel space in order to visualize what that feature stands for. From a theoretical perspective, Bruna et al. [2013] explore the theoretical conditions for a network to be invertible. All of these inverting efforts are instances of the *pre-image problem*, [Kwok and Tsang, 2004].

Bora et al. [2017] have recently showed that it is possible to recover from a trained generator with compressed sensing. Similar to our work, they use a ℓ_2 loss and backpropagate the gradient to the low rank distribution. However, they do not train the generator simultaneously. Jointly learning the representation and training the generator allows us to extend their findings. Santurkar et al. [2017] also use generative models to compress images.

Several works have used an optimization of a latent representation for the express purpose of generating realistic images, e.g. [Portilla and Simoncelli, 2000, Nguyen et al., 2017]. In these works, the total loss function optimized to generate is trained separately from the optimization of the latent representation (in the former, the loss is based on a complex wavelet transform, and in the latter, on separately trained autoencoders and classification convolutional networks). In this work we train the latent representations and the generator together from scratch; and show that at test time we may sample new latent z either with simple parametric distributions or by interpolation in the latent space.

Learning representations. Arguably, the problem of learning representations from data in an unsupervised manner is one of the long-standing problems in machine learning [Bengio et al., 2013, LeCun et al., 2015]. One of the earliest algorithms used to achieve this goal is Principal Component Analysis, or PCA [Pearson, 1901, Jolliffe]. For instance, PCA has been used to learn low-dimensional representations of human faces [Turk and Pentland, 1991], or to produce a hierarchy of features [Chan et al., 2015]. The nonlinear extension of PCA is an autoencoder [Baldi and Hornik, 1989], which is in turn one of the most extended algorithms to learn low-dimensional representations from data. Similar algorithms learn low-dimensional representations of data with certain structure. For instance, in sparse coding [Aharon et al., 2006, Mairal et al., 2008], the representation of one image is the linear combination of a very few elements from a dictionary of features. More recently, Zhang et al. [2016] realized the capability of deep neural networks to map large collections of images to noise vectors, and Bojanowski and Joulin [2017] exploited a similar procedure to learn visual features unsupervisedly. Similarly to us, Bojanowski and Joulin [2017] allow the noise vectors z to move in order to better learn the mapping from images to noise vectors. The proposed GLO is the analogous to these works, in the opposite direction: learn a map *from* noise vectors *to* images. Finally, the idea of mapping between images and noise to learn generative models is a well known technique [Chen and Gopinath, 2000, Laparra et al., 2011, Sohl-Dickstein et al., 2015, Bordes et al., 2017].



Figure 1: Reconstruction results for GLO on LSUN-Bedroom and celebA. From top to bottom, the original images, the reconstruction with a noise representation with $d = 100$ dimensions and $d = 512$ dimensions.

3 Experiments

We organized our experiments as follows. First, Section 3.1 describes the generative models that we compare, along with their implementation details. Section 3.2 reviews the image datasets used in our experiments. Section 3.3 discusses the results of our experiments, including the compression of datasets (Section 3.3.1), the generation of new samples (Section 3.3.2), the interpolation of samples (Section 3.3.3), and the arithmetic of noise vectors (Section 3.3.4).

3.1 Methods

Throughout our experiments, we compare with four different models. The representation space is 100-dimensional for all models.

- **PCA** [Pearson, 1901], equivalent to a linear autoencoder [Baldi and Hornik, 1989], where we retain the top 100 principal components.
- **DCAE** [Masci et al., 2011], a deep convolutional autoencoder minimizing mean-squared error. The encoder f follows the same architecture as the DCGAN discriminator of Radford et al. [2015], after removing the $\text{Tanh}()$ layer, and letting the last layer return 100 feature maps instead of 3. The decoder g follows the same architecture as the DCGAN generator. We do not impose any constraint on the representation space \mathcal{Z} , that is, the output of the encoder. We use the Adam optimizer [Kingma and Ba, 2015] with default parameters.
- **DCGAN** [Radford et al., 2015]. Since GANs do not come with a mechanism (inverse generator) to retrieve the random vector $g^{-1}(x)$ associated with an image x , we estimate this random vector by 1) instantiating a random vector z_0 , and 2) computing updates $z_{i+1} = z_i + \alpha \frac{\partial \ell(g(z_i), x)}{\partial z_i}$ by backpropagation until convergence. Obviously, our experiments

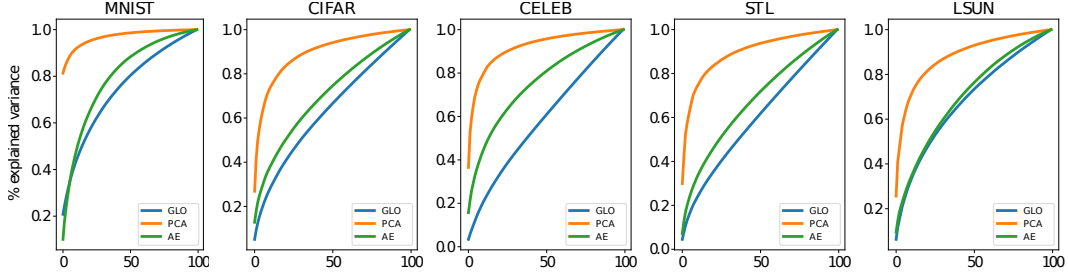


Figure 2: Percent of the representation space explained as a function of the number of singular vectors used to reconstruct it for PCA, AE and GLO. The more flat the curve, the more evenly distributed the energy across the spectrum. Models with flat curves use more of their allotted space to store information.

MSE	MNIST	CIFAR-10	STL-10	CelebA	LSUN
PCA	0.011	0.016	0.039	0.024	0.036
DCAE	0.0002	0.006	0.023	0.015	0.027
GAN	0.016	0.022	0.042	0.030	0.034
GLO (L2)	0.0001	0.005	0.019	0.010	0.021
GLO (LapL1)	0.0001	0.007	0.023	0.012	0.025

Table 1: Reconstruction errors in ℓ_2 loss for all methods and datasets.

measuring reconstruction error are disadvantageous to GANs, since these are models that are *not* trained to minimize this metric. We use the Adam optimizer [Kingma and Ba, 2015] with the parameters suggested by [Radford et al., 2015].

- **GLO (proposed model)**. We will train a GLO model where the generator follows the same architecture as the generator in DCGAN. We use Stochastic Gradient Descent (SGD) to optimize both θ and z , setting the learning rate for θ at 1 and the learning rate of z at 10. We set the representation space $\mathcal{Z} = \mathcal{B}(1, d, 2)$. To ease optimization, we initialize the noise vectors to the solution provided by a PCA.

3.2 Datasets

We evaluate all models on five datasets of natural images. Unless specified otherwise, we use the prescribed training splits to train our generative models. All the images are rescaled to have three channels, 64 pixels on their short side, center-cropped, and normalized to pixel values in $[-1, +1]$.

- **MNIST** [LeCun et al., 1998b] is a set of handwritten digits. We use the original training set, composed of 60,000 grayscale images of size 28×28 pixels.
- **CIFAR-10** [Krizhevsky and Hinton, 2009] is set of images belonging to ten different object categories. We use the original training set, composed of 50,000 color images of size 32×32 pixels.
- **STL-10** [Coates et al., 2011], where we use the set of 100,000 unlabeled images of size 96×96 pixels. Some of these images contain thick black borders.
- **CelebA** [Liu et al., 2015] is a set of 202,599 portraits of celebrities. We use the *aligned and cropped* version, which preprocesses each image to a size of 64×64 pixels.
- **LSUN** [Xiao et al., 2010] is a set of millions of images of scenes belonging to different scene categories. Following the tradition in GAN papers [Radford et al., 2015], we use the 3,033,042 images belonging to the *bedroom* category.



Figure 3: Samples generated using GAN (**top**) and GLO (**bottom**) trained on STL (**left**) and CelebA (**right**). We generate samples by fitting a single Gaussian to the optimal points z .

3.3 Results

In the following, we compare the methods described on Section 3.1 when applied to all the datasets described on Section 3.2. In particular, we evaluate the performance of the methods in the tasks of compressing a dataset, generating new samples, performing sample interpolation, and doing sample arithmetic.

3.3.1 Dataset compression

We start by measuring the reconstruction error in terms of the mean-squared loss $\ell_2(x, x') = \|x - x'\|_2^2$ and the Lap₁ loss (2) Table 1 shows the reconstruction error of all models and datasets for the ℓ_2 and Lap₁ losses. Despite being much simpler than AEs or GANs, GLO obtains better reconstruction error, even when using a Lap₁ loss. Figure 1 shows a few reconstruction examples obtained with different sizes of latent space in GLO.

Figure 2 show the quantity of the representation space explained as a function of the number of eigenvectors used to reconstruct it. Our approach use more efficiently the representation space to spread the information around while PCA concentrates the information in a few directions.

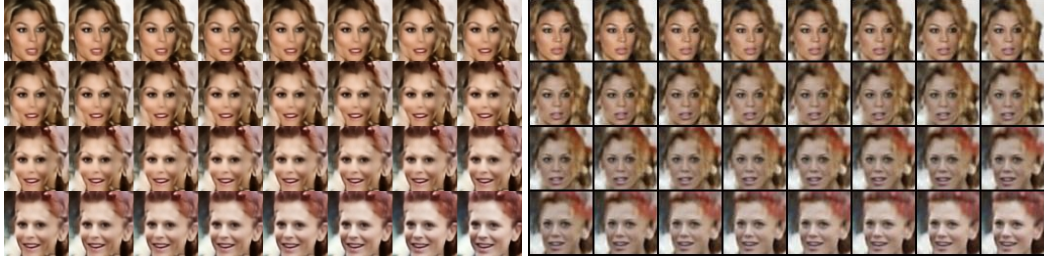


Figure 4: Comparison between GLO trained with a ℓ_2 and Laplacian ℓ_1 loss on interpolation. We use 512 dimensions for z .

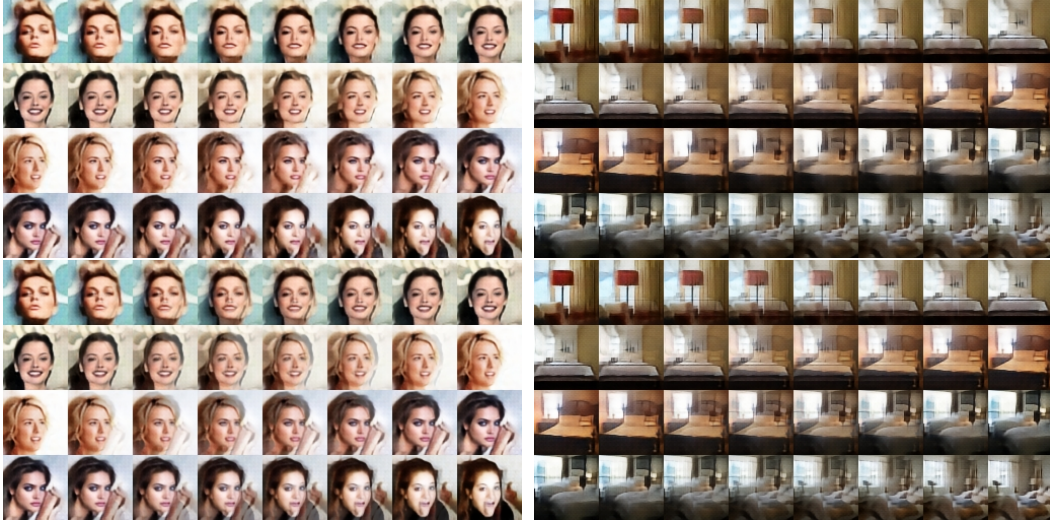


Figure 5: Illustration of image interpolations obtained with GLO for CelebA (**left**) and LSUN (**right**). On top we show the reconstruction that we obtain for interpolated z and compare to a linear mixture (bottom).

3.3.2 Generation of new samples

Figure 3 shows samples from the DCAE, DCGAN, and GLO models for all datasets. In the case of GLO, we fit a Gaussian distribution with full covariance to the representation space, and sample from such distribution to generate new samples. We can see that the samples are quite good, even with this simple model of the Z space; we leave more careful modeling of the Z space for future work.

3.3.3 Sample interpolation

Figure 5 shows a few examples of interpolations between different images from the CelebA dataset and LSUN bedroom. We compare interpolation in the z -space where we linearly interpolate the z and forward them to the model to linear interpolation in the image space. The interpolation in the z are very different from the one in the image space, showing that our generator learns non-linear mapping between the noise and the image. For example, the faces are slowly rotating. The difference between the two type of interpolations is clearer when we are at the middle of the interpolation, that is the images on the 4-th and 5-th columns. The same non-linear transformation is observed on the bedrooms.

Finally, Figure 4 show the difference between the ℓ_2 loss and the Laplacian ℓ_1 loss on high quality model, i.e. a z space of 512 dimensions. The use of sparse loss creates sharper interpolations than the square loss.

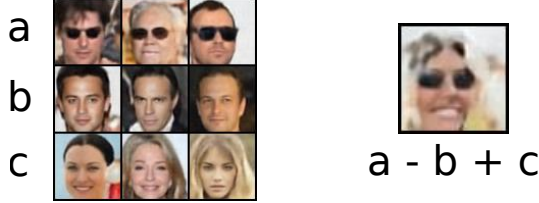


Figure 6: Illustration of feature arithmetic in a GLO model on CelebA. We show that by taking the average hidden representation of row **a**, subtracting the one of row **b** and adding the one of row **c**, we obtain a coherent image.

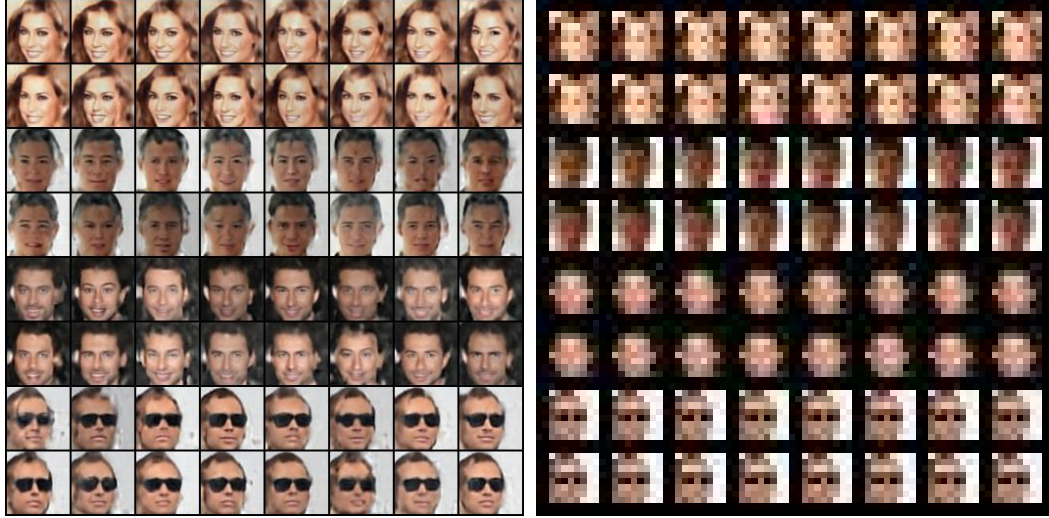


Figure 7: Conditional generation in a GLO model. Each image on the left maps via an 8×8 pooling operator to the corresponding image on the right. The model is trained as normal, but at sampling time, each z variable is initialized independently, and we measure the loss on the targets after the pooling. Note the variety of hallucinations generated. Although the conditioning is not perfect (there are slight variations on the right side image), it is quite good; the average square error of each projected generated image from its projected target is about .0002 of the square norm of the projected target.

3.3.4 Noise vector arithmetic

In the spirit of Radford et al. [2015], we showcase the effect of simple arithmetic operations in the noise space of a GLO model. More precisely, we average the noise vector of three images of men wearing sunglasses, remove the average noise vector of three images of men not wearing sunglasses, and add the average noise vector of three images of women not wearing sunglasses. The resulting image resembles a woman wearing sunglasses glasses, as shown in Figure 6.

3.3.5 Conditional generation

Our model can also be used for conditional generation. To do this, we take a fully trained model, and measure the loss only on the conditioned variables. For example, to hallucinate 8×8 upsamplings of an 8×8 image (i.e. condition on the output of an 8×8 pooling operation), when backpropagating to find the z variable, we only measure error on the output of the pooling operation applied to the generated image. That is, if x is the small image to condition on, and P is the pooling operator that takes a large image and maps it down to the small image, we minimize $\|P(g(z)) - x\|^2$ instead of $\|g(z) - x\|^2$. In a similar way, we can condition on the right side pixels matching the target but the left side being free, etc. Because the optimization over z does not usually lead to 0 error, the conditioning is not exact; but the error can often be made quite small.

4 Discussion

The experimental results presented in this work suggests that, in the image domain, we can recover many of the properties of generative models trained with GANs with convnets trained with reconstruction losses. While this does not invalidate the promise of GANs as generic models of uncertainty or as methods for building generative models, it does suggest that in order to more fully test the adversarial construction, researchers will need to move beyond images and convnets. On the other hand, practitioners who care only about generating images for a particular application, and find that the parameterized discriminator does improve their results can use reconstruction losses in their model searches, alleviating some of the instability of GAN training.

While the results are promising, they are not yet to the level of the latest results obtained by recent works that generate images larger than the 64×64 images generated here. There is much work to be done here to fill that gap. In particular, exploring other loss functions and model architectures, adding more structure to the Z space, and using more sophisticated sampling methods after training the model should all improve the results obtained by our method.

Acknowledgments. Thanks to Mark Tygert and Yann LeCun for planting the seeds that grew into this work.

References

- M. Aharon, M. Elad, and A. Bruckstein. *rmk-svd: An algorithm for designing overcomplete dictionaries for sparse representation*. *IEEE Transactions on signal processing*, 2006.
- F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Convex optimization with sparsity-inducing norms. In *Optimization for Machine Learning*. 2011.
- P. Baldi and K. Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks*, 1989.
- Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 2013.
- P. Bojanowski and A. Joulin. Unsupervised Learning by Predicting Noise. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2017.
- A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed Sensing using Generative Models. *ArXiv e-prints*, Mar. 2017.
- F. Bordes, S. Honari, and P. Vincent. Learning to Generate Samples from Noise through Infusion Training. *ArXiv e-prints*, Mar. 2017.
- H. Bourlard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 1988.
- J. Bruna, A. Szlam, and Y. LeCun. Signal recovery from pooling representations. *arXiv preprint arXiv:1311.4025*, 2013.
- T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma. PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Transactions on Image Processing*, Dec. 2015.
- S. S. Chen and R. A. Gopinath. Gaussianization. In *Proceedings of the 13th International Conference on Neural Information Processing Systems*. MIT Press, 2000.
- A. Coates, A. Ng, and H. Lee. An analysis of single-layer networks in unsupervised feature learning. In *AISTATS*, 2011.
- A. Creswell and A. A. Bharath. Inverting The Generator Of A Generative Adversarial Network. *ArXiv e-prints*, Nov. 2016.
- E. L. Denton, S. Chintala, R. Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, 2015.

- J. Donahue, P. Krähenbühl, and T. Darrell. Adversarial feature learning. *arXiv preprint arXiv:1605.09782*, 2016.
- V. Dumoulin, I. Belghazi, B. Poole, A. Lamb, M. Arjovsky, O. Mastropietro, and A. Courville. Adversarially learned inference. *arXiv preprint arXiv:1606.00704*, 2016.
- I. Goodfellow. NIPS 2016 Tutorial: Generative Adversarial Networks. *ArXiv e-prints*, Dec. 2017.
- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. 2014.
- G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 2006.
- S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and Locally Consistent Image Completion. *ACM Transactions on Graphics (Proc. of SIGGRAPH 2017)*, 36(4):107:1–107:14, 2017.
- I. Jolliffe. *Principal component analysis*. Wiley Online Library.
- D. Kingma and J. Ba. Adam: A method for stochastic optimization. *ICLR*, 2015.
- D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images. 2009.
- J.-Y. Kwok and I.-H. Tsang. The pre-image problem in kernel methods. *IEEE transactions on neural networks*, 2004.
- V. Laparra, G. Camps-Valls, and J. Malo. Iterative gaussianization: from ica to random rotations. *IEEE transactions on neural networks*, 2011.
- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998a.
- Y. LeCun, C. Cortes, and C. J. C. Burges. The MNIST database of handwritten digits, 1998b. URL <http://yann.lecun.com/exdb/mnist/>.
- Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 2015.
- C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.
- H. Ling and K. Okada. Diffusion distance for histogram comparison. In *Computer Vision and Pattern Recognition*, 2006.
- Z. C. Lipton and S. Tripathi. Precise Recovery of Latent Vectors from Generative Adversarial Networks. *ArXiv e-prints*, Feb. 2017.
- Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *ICCV*, 2015.
- J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 2008.
- J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. *ICANN*, 2011.
- T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- A. Nguyen, J. Yosinski, Y. Bengio, A. Dosovitskiy, and J. Clune. Plug & play generative networks: Conditional iterative generation of images in latent space. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

- K. Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901.
- J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vision*, 40(1):49–70, Oct. 2000.
- A. Radford, L. Metz, and S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ArXiv e-prints*, Nov. 2015.
- T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, 2016.
- S. Santurkar, D. Budden, and N. Shavit. Generative compression. *arXiv*, 2017.
- J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. *arXiv preprint arXiv:1503.03585*, 2015.
- M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991.
- D. Ulyanov, A. Vedaldi, and V. Lempitsky. Adversarial Generator-Encoder Networks. *ArXiv e-prints*, Apr. 2017.
- P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*. ACM, 2008.
- J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, 2010.
- M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*. Springer, 2014.
- C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning requires rethinking generalization. *ArXiv e-prints*, Nov. 2016.
- J. Zhao, M. Mathieu, and Y. LeCun. Energy-based generative adversarial network. *arXiv preprint arXiv:1609.03126*, 2016.
- J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros. Generative visual manipulation on the natural image manifold. In *ECCV*, 2016.