
Generative Imaging and Image Processing via Generative Encoder

Anonymous Author(s)

Affiliation
Address
email

Abstract

1 This paper introduces a novel generative encoder (GE) model for generative imaging
2 and image processing with applications in compressed sensing and imaging,
3 image compression, denoising, inpainting, deblurring, and super-resolution. The
4 GE model consists of a pre-training phase and a solving phase. In the pre-training
5 phase, we separately train two deep neural networks: a generative adversarial
6 network (GAN) with a generator \mathcal{G} that captures the data distribution of a given
7 image set, and an auto-encoder (AE) network with an encoder \mathcal{EN} that compresses
8 images following the estimated distribution by GAN. In the solving phase, given a
9 noisy image $x = \mathcal{P}(x^*)$, where x^* is the target unknown image, \mathcal{P} is an operator
10 adding an additive, or multiplicative, or convolutional noise, or equivalently given
11 such an image x in the compressed domain, i.e., given $m = \mathcal{EN}(x)$, we solve the
12 optimization problem

$$z^* = \operatorname{argmin}_z \|\mathcal{EN}(\mathcal{G}(z)) - m\|_2^2 + \lambda \|z\|_2^2$$

13 to recover the image x^* in a generative way via $\hat{x} := \mathcal{G}(z^*) \approx x^*$, where $\lambda > 0$ is
14 a hyperparameter. The GE model unifies the generative capacity of GANs and the
15 stability of AEs in an optimization framework above instead of stacking GANs and
16 AEs into a single network or combining their loss functions into one as in existing
17 literature. Numerical experiments show that the proposed model outperforms
18 several state-of-the-art algorithms.

19

1 Introduction

20 Recently, deep learning-based structures have become an effective tool for imaging and image
21 processing with compressed or compressible information. One stream of such studies is an end-to-
22 end training with a deep neural network (DNN) mapping a source image to a reconstructed image with
23 desired properties. Due to the powerful representation capacity of DNNs, DNNs can approximate
24 the desired imaging or image processing procedure well as long as training data are sufficiently
25 good. To ease the training of DNNs, special NN structures are proposed, e.g., DNNs that mimic a
26 traditional optimization algorithms for imaging or image processing (Adler et al. [2017], Yang et al.
27 [2016], Ledig et al. [2017], Lu et al. [2018]), autoencoders (AEs) with built-in image compression
28 and reconstruction (Vincent et al. [2010], Schlemper et al. [2017]). The dimension reduction in AEs
29 plays a key role to enhance the performance of DNNs as important as sparsity in traditional imaging
30 and image processing algorithms.

31 Especially, deep convolutional encoders, $\mathcal{EN}(x; \theta_{EN})$ with parameters θ_{EN} , can adaptively capture
32 the low-dimensional structure of a source image x through repeated applications of convolution,
33 pooling, and nonlinear activation functions, and finally outputs a small feature vector z ; and deep
34 convolutional decoders, $\mathcal{DE}(z; \theta_{DE})$ with parameters θ_{DE} , can efficiently reconstruct the image x .

35 via a sequence of deconvolution, up-sampling, and nonlinear activation functions acting on z (see
 36 Figure 1 for a visualization of the structure of an autoencoder). Parameters in the encoder and decoder
 37 are jointly tuned such that the input and output of the autoencoder match well via the following
 38 optimization

$$\min_{\theta_{EN}, \theta_{DE}} \mathbb{E}_{x \sim p_{data}(x)} [\|\mathcal{D}\mathcal{E}(\mathcal{EN}(x; \theta_{EN}); \theta_{DE}) - x\|_2^2], \quad (1)$$

39 where p_{data} is the image data distribution. Armed with the powerful representation capacity of
 40 DNNs, deep convolutional autoencoders are capable of learning nonlinear transforms that create
 41 highly sparse or very low-dimensional representations of images from a certain distribution, while
 42 keeping the accuracy of image reconstruction. Due to the least square nature of (1), AEs penalizes
 43 pixel-wise error and hence prefers smooth images that lack fine details of the original image and
 44 avoids generative changes in image reconstruction.

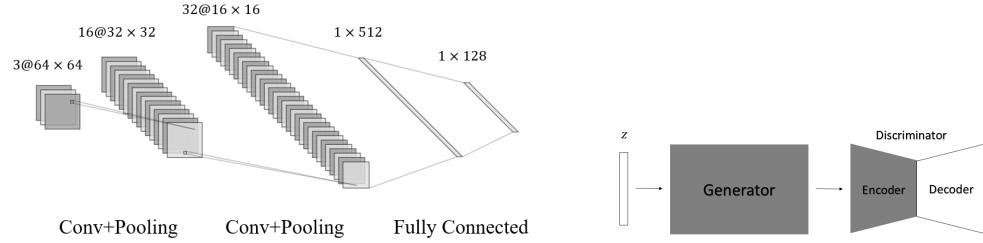


Figure 1: Typical encoding network in convolutional autoencoder.

Figure 2: BEGAN structure.

45 Another stream of deep learning approaches (Warde-Farley and Bengio [2017], Bora et al. [2017], Yeh
 46 et al. [2017], Yan and Wang [2017], Kupyn et al. [2018], Chen et al. [2018]) is based on generative
 47 models, e.g., the generative adversarial network (GAN) (Goodfellow et al. [2014]) and its variants
 48 (Radford et al. [2016], Arjovsky et al. [2017], Zhao et al. [2017], Berthelot et al. [2017]). A GAN
 49 consists of a generator \mathcal{G} and a discriminator \mathcal{D} that are trained through an adversarial procedure (see
 50 Figure 2 for a visualization of the structure of a GAN). The generator, \mathcal{G} , takes a random vector z from
 51 a given distribution p_z and outputs a synthetic sample $\mathcal{G}(z; \theta_G)$, where θ_G refers to the corresponding
 52 parameters of \mathcal{G} . The discriminator, \mathcal{D} with parameters θ_D , takes an input x and outputs a value
 53 $\mathcal{D}(x; \theta_D) \in [0, 1]$ denoting the probability of the input following the data distribution p_{data} . \mathcal{G} is
 54 trained to generate samples to fool \mathcal{D} into thinking that the generated sample is real, and \mathcal{D} is learned
 55 to distinguish between real samples from the data distribution versus synthetic fake data from \mathcal{G} via
 56 an adversarial game below:

$$\min_{\theta_G} \max_{\theta_D} V(\theta_D, \theta_G) = \mathbb{E}_{x \sim p_{data}(x)} [\log \mathcal{D}(x; \theta_D)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - \mathcal{D}(\mathcal{G}(z; \theta_G); \theta_D))]. \quad (2)$$

57 The discriminator \mathcal{D} in the adversarial learning above can be applied to enhance image quality
 58 as in (Warde-Farley and Bengio [2017]). The generator \mathcal{G} can also be applied as a tool for data
 59 augmentation (Bowles et al. [2018], Huang et al. [2018]) or as an inverse operator that returns the
 60 desired image from compressed measurements in compressed sensing (Bora et al. [2017]). However,
 61 solving the optimization in (2) or its variants is challenging and the solution might not be stable, e.g.,
 62 the generator might create images with unreasonable global content, although fine details of image
 63 content are better than those generated by AEs.

64 In this paper, we introduce a novel generative encoder (GE) model that takes advantage of AEs and
 65 GANs simultaneously for generative imaging and image processing with applications in compressed
 66 sensing and imaging, image compression, denoising, inpainting, deblurring, and super-resolution.
 67 The GE model consists of a pre-training phase and a solving phase. In the pre-training phase, we
 68 separately train a GAN and an AE. The generator \mathcal{G} in GAN can capture the data distribution of a
 69 given image set and hence can generate training data to enhance the encoder \mathcal{EN} of the AE such that
 70 \mathcal{EN} can better compress images following the target data distribution. In the solving phase, given a
 71 noisy image $x = \mathcal{P}(x^*)$, where x^* is the target unknown image, \mathcal{P} is an operator adding an additive,
 72 or multiplicative, or convolutional noise, or equivalently given such an image x in the compressed
 73 domain, i.e., given $m = \mathcal{EN}(x)$, we solve the optimization problem

$$z^* = \operatorname{argmin}_z \|\mathcal{EN}(S(\mathcal{G}(z))) - m\|_2^2 + \lambda \|z\|_2^2 \quad (3)$$

74 to recover the image x^* in a generative way via $\hat{x} := \mathcal{G}(z^*) \approx x^*$, where S is a down-sample or an
 75 up-sample operator to balance the dimension of the output of \mathcal{G} and the input of \mathcal{EN} , and $\lambda > 0$ is
 76 a hyperparameter. The GE model unifies the generative capacity of GANs and the stability of AEs
 77 in an optimization framework in (3) instead of stacking GANs and AEs into a single network or
 78 combining their loss functions into one as in existing literature.

79 The novelty and the advantages of the proposed GE model can be summarized as follows.

- 80 1. Different to existing methods, the training of GAN and AE in the GE model is separate to
 81 reduce the competition of GAN and AE to maximize the generalization capacity of GAN
 82 and the dimension reduction ability of AE. Numerical results will show that GE outperforms
 83 traditional methods, end-to-end convolutional DNN models, and previous GAN-based
 84 models achieving a better compression ratio.
- 85 2. The training data of AE is augmented by the generator of GAN such that the dimension
 86 reduction of AE can adapt to the target data distribution, making the AE more compatible
 87 with GAN and increasing the generalization flexibility of AE.
- 88 3. We only unify the most attractive parts of GAN and AE, i.e., the generator of GAN for
 89 generalization and the encoder of AE for data-driven compression, which is different to
 90 existing works that combine all components of GAN and AE together in a single network.
- 91 4. Instead of creating an end-to-end neural network, a regularized least square problem (3) is
 92 proposed to search for the best reconstruction that fits data measurements in the compressed
 93 domain. On the one hand, the reconstruction has been stabilized via reducing the search
 94 domain from the image domain to the compressed domain, e.g., the encoder can rule out
 95 extreme results generalized by the generator. On the other hand, the reconstruction still
 96 inherits the generalization capacity to enhance the detailed reconstruction which is often
 97 lost in traditional AEs.
- 98 5. The GE model is a general framework with various applications such as denoising, deblurring,
 99 super-resolution, inpainting, compressed sensing, etc. Numerical experiments show
 100 that the proposed model can generally produce competitive or better outcomes compared to
 101 existing methods.

102 2 Applications

103 Let's briefly introduce the applications of the GE model in this paper.

104 **Compressed Sensing** Given a measurement vector $y = Ax^* + \epsilon$, where A is a sensing matrix
 105 satisfying the restricted isometry property (RIP), ϵ is a noise vector, the compressed sensing problem
 106 seeks to recover x^* . If x^* is sparse, the recovery via an ℓ^1 -penalized least square problem is
 107 guaranteed by the compressed sensing theory in (Candes et al. [2006], Donoho [2006]). The RIP
 108 condition is satisfied when A is a random Gaussian matrix, and natural images are generally sparse
 109 after an appropriate transform, e.g., wavelet transforms. Therefore, compressed sensing has been a
 110 successful tool in imaging science.

111 Recently, pioneer works including (Bora et al. [2017]) have explored the application of generative
 112 models to improve traditional compressed sensing algorithms. The main idea is to apply the generator
 113 \mathcal{G} to generate a synthesized image $\mathcal{G}(z)$ for z in a compressed space. Bora et al. proposed

$$z^* = \arg \min_z \|A\mathcal{G}(z) - y\|_2^2 + \lambda \|z\|_2^2 \quad (4)$$

114 to find the reconstruction image $\mathcal{G}(z^*) \approx x^*$. They also provided a theoretical guarantee on the relief
 115 of sparsity condition on the compressed space and thus serves as a nice benchmark for our model.

116 **Denoising and Inpainting** Denoising and inpainting have the same problem statement in mathemat-
 117 ics. Given a measurement $y = x^* + \epsilon$ or $y = x^* \circ \epsilon$, where ϵ is a certain random or structured noise
 118 and \circ represents the Hadamard product, denoising and inpainting seek to recover x^* from y .

119 Traditional denoising or inpainting techniques generally rely on the sparsity of x^* after an approximate
 120 transform in a certain metric, e.g., KSVD (Aharon et al. [2006]), GSR (Zhang et al. [2014]), BM3D
 121 (Dabov et al. [2009]), NLM (Buades et al. [2005]), and total variation (TV) regularization (Getreuer
 122 [2012]). Deep learning approaches have become more popular than traditional methods recently, e.g.,

123 deep autoencode methods (Vincent et al. [2010], Pathak et al. [2016], Yu et al. [2018]), especially
 124 when the hidden information of noisy or damaged images is not visually obviously.

125 **Deblurring** Blurring an image is commonly modeled as the convolution of a point-spread function
 126 over an original sharp image. Deblurring aims to reverse this process. Mathematically speaking,
 127 given a measurement $y = x * h$, where h is an unknown convolution kernel function and $*$ represents
 128 the convolution operator. Deblurring seeks to recover x with certain assumptions on h and x to ease
 129 the ill-posedness. Sparse coding (Dong et al. [2011]) and kernel estimation (Xu and Jia [2010]) are
 130 effective methods for image deblurring. Deep convolutional networks have also been applied to this
 131 problem recently (Yan and Shao [2016]), especially with the help of GAN (Berthelot et al. [2017]).

132 **Super-resolution** Super-resolution imaging aims at generating high-resolution images from low-
 133 resolution ones. For example, given a measurement $y = S(x^*)$, where S is a down-sampling operator,
 134 or a convolution operator with a convolution kernel function decaying quickly in the Fourier domain.
 135 Super-resolution imaging seeks to recover x^* with certain assumptions on S and x^* to ease the
 136 ill-posedness. Traditionally, Bicubic interpolation and sparse-coding (Dong et al. [2011]) are popular
 137 tools to increase image resolution. Recently, deep convolutional neural networks have also been
 138 applied to solve this problem with great success (Dmitry Ulyanov [2017], Ledig et al. [2017]).

139 3 Implementation of the Proposed Framework

140 Let's introduce the detailed implementation of the GE model in this section. In this paper, we apply
 141 the BEGAN (Berthelot et al. [2017]) and the convolutional AE in the implementation. In fact, our
 142 framework is broadly applicable to various GANs and AEs for the customized settings. The overall
 143 training procedure of the GE model is summarized in Algorithm 3 and visualized in Figure 3.

Algorithm 1 GE: generative encoder model.

1. Pre-train a generator using any GAN.
 2. Pre-train a convolutional AE using both real and fake images (generated by GAN).
 3. Take the generator \mathcal{G} of GAN and the encoder \mathcal{EN} of AE to form the generative encoder.
 4. Given the measurement $m = \mathcal{EN}(x^*)$. Find $z^* = \arg \min_z \|\mathcal{EN}(S(\mathcal{G}(z))) - m\|_2^2 + \lambda \|z\|_2^2$ and return $\hat{x} = \mathcal{G}(z^*)$.
-

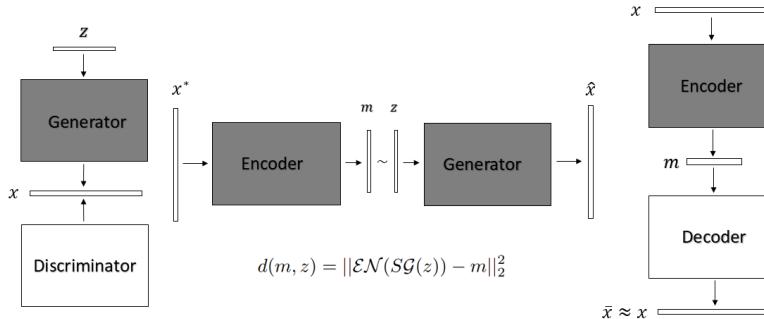


Figure 3: GE: Generative Encoder. Left: pre-trained generative adversarial network; middle: GE sensing and reconstructing flow; right: pre-trained convolutional autoencoder

144 The key idea of the GE model is to combine the generator \mathcal{G} and the encoder \mathcal{EN} from a GAN and an
 145 AE model, respectively, to take the advantages of these two models via a new optimization framework
 146 in (3). The term $d(m, z) = \|\mathcal{EN}(S(\mathcal{G}(z))) - m\|_2^2$ itself could serve as the loss function of the
 147 GE model. We introduce the ℓ^2 term to regularize the highly nonconvex function $d(m, z)$. Besides,
 148 looking for a solution with a small ℓ^2 -norm also agrees with the GAN model that $z \sim \mathcal{N}(0, I)$, i.e., z
 149 has a higher probability to have a smaller magnitude. An interesting extension would be replacing
 150 the ℓ^2 regularization with a DNN for a data-driven regularization. This is left as a future work.

151 **3.1 Compressed Sensing**

Suppose the original image is x^* and we are able to design a sensor \mathcal{EN} . Given the measurement $m = \mathcal{EN}(x^*)$, in order to find the reconstruction \hat{x} that is as close to x^* as possible, we find z in the latent space such that the compressed measurement of its generated image, i.e., $\mathcal{EN}(\mathcal{G}(z))$, is as close to the measurement of the real image $\mathcal{EN}(x^*)$ as possible. Therefore, we solve

$$z^* = \arg \min_z \|\mathcal{EN}(\mathcal{G}(z)) - m\|_2^2 + \lambda \|z\|_2^2$$

152 to identify $\hat{x} := \mathcal{G}(z^*) \approx x^*$ as the reconstruction.

153 **3.2 Denoising, Deblurring, Super-Resolution & Inpainting**

154 The solution of the denoising, deblurring, super-resolution, and inpainting can be obtained by solving

$$z^* = \arg \min_z \|\mathcal{EN}(SG(z)) - \mathcal{EN}(x^\dagger)\|_2^2 + \lambda \|z\|_2^2, \quad (5)$$

156 where x^\dagger is the given noisy image, or the image with missing pixels, or the blurry image, or the
157 image of low-resolution, constructed from an unknown target image x^* . Then the reconstructed
158 image is set as $\hat{x} = \mathcal{G}(z^*) \approx x^*$. In (5), S is an adjustment operator, which is an identity for
159 denoising and deblurring, a masking operator for inpainting, and a dimension adjustment operator for
160 super-resolution.

161 **4 Training Details & Empirical Results**

162 In this section, we describe training details and present experimental results to support the GE model.

163 **4.1 Training Details**

164 **Training Data:** We use the CelebA dataset (Liu et al. [2015]) containing more than 200,000 celebrity
165 images cropped to sizes of $64 \times 64 \times 3$ and $128 \times 128 \times 3$, respectively.

166 **GAN:** The BEGAN is trained with the data sets above and the corresponding generator \mathcal{G} is adopted
167 in our GE model. The discriminator of the BEGAN is, in fact, an AE (See Figure 2 for the structure
168 of the BEGAN). For $64 \times 64 \times 3$ and 64 for $128 \times 128 \times 3$ datasets, the input dimensions of \mathcal{G}
169 are 128 and 64 , respectively; the numbers of convolutional layers of the encoder in BEGAN are 12
170 and 15 , respectively; the numbers of convolutional layers of \mathcal{G} are 9 and 11 , respectively, with one
171 up-sampling every two convolutions.

172 **AE:** We test two different types of AEs and the corresponding GEs are denoted as $GE0$ and $GE1$.
173 In the first type, we use the encoding part of BEGAN’s discriminator as the encoder in GE. Due to
174 the special encoding-decoding structure, the discriminator of BEGAN is trained to compress real
175 and fake images in the same manner as an autoencoder. In the second type, we adopt convolutional
176 autoencoders and train them separately with both real images and fake images generated by BEGAN.

177 We use the following notation to denote the specific structure of the \mathcal{EN} in $GE0$ and $GE1$. For
178 example, $GE1(d=4, f=16, m=128)$ indicates that: 1) the encoder \mathcal{EN} in our GE model has 4
179 convolutional layers; 2) the output dimension of \mathcal{EN} is $m=128$; 3) in the n -th convolutional layer,
180 the number of filters is $n \times 16$ for $GE1$, and $\lceil \frac{n}{3} \rceil \times 16$ for $GE0$.

181 **Optimization in GE:** ADAM optimizer (Kingma and Ba [2014]) is used with a learning rate 0.1 to
182 solve (3). For images of size $64 \times 64 \times 3$, we use 500 iterations with 2 random starts to find the
183 optimal reconstruction; for larger images, 700 iterations with 2 random starts are used.

184 **Baseline Methods:** We compare GE with two baseline methods under the same conditions:

- 185 • **Lasso:** Let x^* be an unknown image and it has a sparse representation in a dictionary Ψ .
186 Given $b = A\Psi^\dagger x^*$, where \dagger denotes the pesudo-inverse and A is a random Gaussian matrix,
187 Lasso solves the minimization problem: $\beta^* = \arg \min_\beta \|A\beta - b\|_2^2 + \alpha \|\beta\|_1$ to estimate
188 x^* by $\hat{x} := \Psi\beta^*$. We choose $\alpha = 0.1$ and Ψ as an overcomplete discrete cosine dictionary
189 for all experiments.

190

- 191 • *GA*: This benchmark uses the model in (4) proposed by (Bora et al. [2017]). Instead of
 192 DCGAN, which is used in the original work, we use BEGAN’s generator to make the
 193 comparison consistent since the GE model uses BEGAN.

194 **4.2 Experimental Results**

195 **Reconstruction - $64 \times 64 \times 3$ Images** Figure 4 presents the comparison of *GE0*, *GE1*, and *GA*
 196 with a measurement size $m = 128$ for $64 \times 64 \times 3$ images. The measurement size for *Lasso* is set
 197 to 500 because *Lasso* totally fails when $m = 128$.

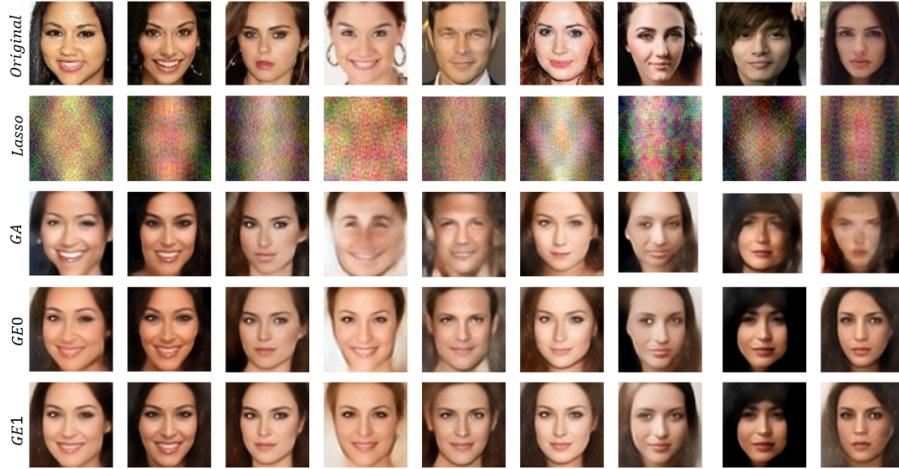


Figure 4: Reconstruction results on CelebA dataset $64 \times 64 \times 3$ images with 128 measurements.
 1st row: original image; 2nd row: *Lasso* ($m = 500$); 3rd row: *GA*; 4th row: *GE0* ($d = 12, f = 64, m = 128$); 5th row: *GE1* ($d = 4, f = 16, m = 128$)

198 According to visual inspection, GE model is more stable than *GA* and *Lasso* and the GE model can
 199 also preserve most fine details of faces as shown in Figure 4. Actually, the reconstruction of *GA*
 200 using DCGAN in (Bora et al. [2017]) is comparable to our *GA* reconstruction using BEGAN. Hence,
 201 the GE model outperforms *GA* and *Lasso*.

202 Besides, two proposed frameworks, *GE0* and *GE1*, have high compatibility and detail-retaining
 203 properties. Note that the network size of *GE0* is much larger than that of *GE1*. Hence, we have
 204 shown that GE model is not sensitive to network size and is very efficient.

	Lasso	GA	GE0
MSE	0.024	0.015	0.009

Table 1: Reconstruction loss.

	Fake	Real	Ratio
MSE	0.00036	0.00928	0.039

Table 2: MSE for real & fake images.

205 To make a quantitative comparison, the sampled average of the mean square error (MSE) between
 206 restored and original images is computed for *GE0* and *GA* in Table 1, which shows that *GE0*
 207 outperforms *GA* by 50%. In addition, we notice that the variation of the MSE of *GE* generally is
 208 smaller *GA*, which also indicates that *GE* is more stable than *GA*.

209 The reconstruction error comes from two sources: a systematic error caused by the compression and
 210 reconstruction mechanism, and a representation error caused by the gap between generated and real
 211 images. The reconstruction loss of real and fake images is shown in Table 2. The reconstruction for
 212 fake images generated by the generator only has one source of error: the systematic error. The result
 213 shows that more than 95% of the current MSE comes from the gap between the generated space and
 214 real pictures. Therefore, if training data are better and the GAN design is better, the MSE can be
 215 further reduced and the GE model becomes better.

216 Figure 5 visually demonstrates how measurement size affects the quality of reconstruction. Generally
 217 speaking, restoration with a larger measurement size has better quality, but the bound of the recon-

218 struction error is still determined by the capacity of the generator. In addition, the reconstruction
 219 under measurement $m = 64$ is still reasonable, although the restoration tends to be slightly distorted,
 220 which indicates that the GE model could be applied to significantly boost sensing efficiency. Figure 6
 221 quantifies the influence of the measurement size on the MSE of image reconstruction.



Figure 5: Reconstruction with increasing measurement size.
 1st row: original image; 2nd row: $GE1(d = 4, f = 16, m = 64)$; 3rd row: $GE1(d = 4, f = 16, m = 128)$, 4th row:
 $GE1(d = 4, f = 16, m = 256)$

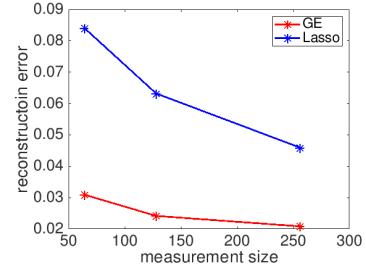


Figure 6: $GE1$ reconstruction loss with different measurement sizes.

222 Finally, we would like to highlight that it is possible to enhance the encoder by sacrificing the decoder
 223 when training the autoencoder, as the decoder is not used anymore after training. By limiting the depth
 224 and width of the decoder, we can force the encoder to be more informative. Moreover, in different
 225 applications, customized encoders can be trained to cater for different data sets and challenges.

226 **Reconstruction - $128 \times 128 \times 3$ Images** Figure 7 shows a comparison between $GE0$ and GA with
 227 a measurement size $m = 64$ for $128 \times 128 \times 3$ images with a measurement rate $\rho = 0.0013$, which is
 228 very attractive for image compression. This measurement rate is mainly attributed to the quality of
 229 the pre-trained generator as well as the smoothness of large images. BEGAN itself has a very good
 230 generation quality for $128 \times 128 \times 3$ images and consequently, GE is able to deliver better results.



Figure 7: Reconstruction results on CelebA dataset $128 \times 128 \times 3$ images with 64 measurements.
 1st row: original image; 2nd row: GA ; 3rd row: $GE0(d = 15, f = 64, m = 64)$

231 **Denoising** We add Gaussian random noise with standard deviation of $\sigma = 0.4$ to pictures to generate
 232 noisy images. Suppose the noisy image is x^\dagger , the GE model searches for a synthesized image that
 233 matches the input noisy image best in the compressed domain by solving

$$z^* = \arg \min_z \|\mathcal{E}\mathcal{N}(SG(z)) - \mathcal{E}\mathcal{N}(x^\dagger)\|_2^2 + \lambda \|z\|_2^2. \quad (6)$$

234 The final reconstruction is given by $\hat{x} = \mathcal{G}(z^*)$. As Figure 8 reflects, benefited from the high
 235 compression rate due to the data-driven encoder of GE, GE is quite robust to noise, while BM3D
 236 cannot recover clear images when the GE model still works.

237 **Deblurring** To obtain a blurred image, rotationally symmetric Gaussian lowpass filter is applied to
 238 the clear images. To reverse the convolution process, blind deconvolution can be done under the same
 239 scheme as denoising. Suppose x^\dagger is the blurred image. The same algorithm as in (6) is applied to
 240 reconstruct a clear image in the GE model. For comparison, a blind deconvolution based on point
 241 spread function (PSF) (Lam and Goodman [2000]) is listed together with GE in Figure 9. Compared
 242 with PSF, GE provides finer details on the resulting images. Again, the main source of error in the

243 restoration by GE is the gap between generation and the original image. A well-trained GAN with
 244 sufficiently good training data can improve the performance of GE.

245 **Super-resolution** We downsample original $64 \times 64 \times 3$ images to $16 \times 16 \times 3$. The performance
 246 of Bicubic interpolation (Ruangsang and Aramvith [2017]) is included as a benchmark. Instead
 247 of inputting the downsampled image into GE directly, we use the image constructed by Bicubic
 248 interpolation as input x^\dagger in (6), which is equivalent to use the Bicubic interpolation as a preconditioner
 249 of the optimization in (6). Figure 10 shows the results of the Bicubic interpolation and the GE model.

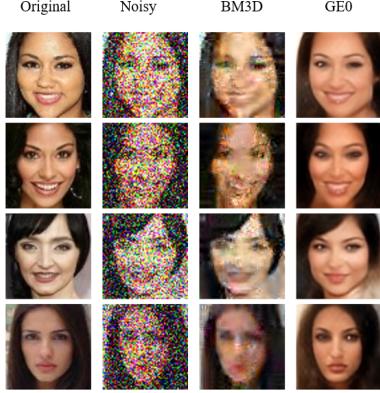


Figure 8: Blind Denoising.

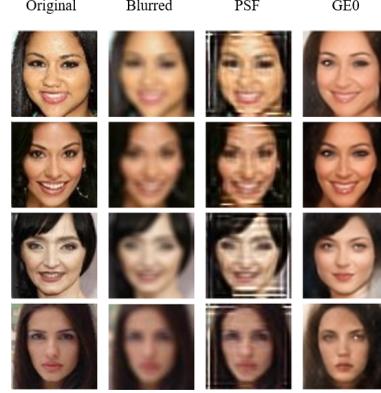


Figure 9: Blind Deblurring.

250 **Inpainting** The first group of images is partially masked by a 30×30 block that sets the value in the
 251 missing area of the image 0, and the second group is covered by hand-drawn lines and text. Suppose
 252 Ω is the mask area and M is a binary matrix taking value 0 at a position in Ω and value 1 outside
 253 Ω . Let S be a masking operator such that $S(x) = x \circ M$, where \circ is the Hadamard product. Again,
 254 assume x^\dagger is the corrupted input and reconstruct the inpainted images via (6) with the masking
 operator S . Reconstructed images are shown in Figure 11.

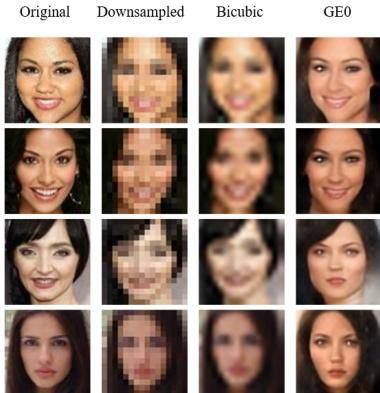


Figure 10: Super-resolution.

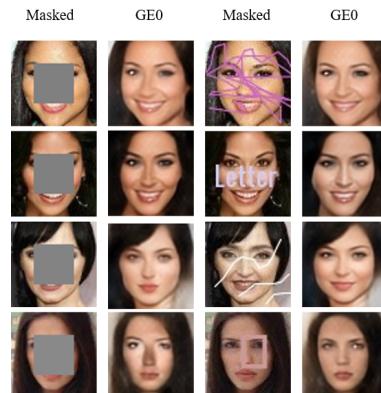


Figure 11: Inpainting.

255

256 5 Conclusion

257 In this paper, we have introduced the generative encoder (GE) model, an effective yet flexible
 258 framework that produces promising outcomes for generative imaging and image processing with broad
 259 applications including compressed sensing, denoising, deblurring, super-resolution, and inpainting.
 260 The GE model unifies GANs and AEs in an innovative manner that can maximize the generative
 261 capacity of GANs and the compression ability of AEs to stabilize image reconstruction, leading to
 262 promising numerical performance that outperforms several state-of-the-art algorithms. The code of
 263 this paper will be available in the authors' personal homepages.

264 **References**

- 265 A. Adler, D. Boublil, and M. Zibulevsky. A deep learning approach to block-based compressed sensing of
266 images. *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct 2017. ISSN
267 2473-3628. doi: 10.1109/MMSP.2017.8122281.
- 268 M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse
269 representation. *IEEE Transactions on Signal Processing*, 54(11):4311 – 4322, Oct 2006. ISSN 1053-587X.
270 doi: 10.1109/TSP.2006.881199.
- 271 M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In D. Precup and
272 Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70
273 of *Proceedings of Machine Learning Research*, pages 214–223, International Convention Centre, Sydney,
274 Australia, 06–11 Aug 2017. PMLR. URL <http://proceedings.mlr.press/v70/arjovsky17a.html>.
- 275 D. Berthelot, T. Schumm, and L. Metz. BEGAN: boundary equilibrium generative adversarial networks. *CoRR*,
276 abs/1703.10717, 2017. URL <http://arxiv.org/abs/1703.10717>.
- 277 A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed sensing using generative models. *ICML'17*
278 Proceedings of the 34th International Conference on Machine Learning, 70:537–546, Aug 2017. ISSN
279 2640-3498.
- 280 C. Bowles, L. J. Chen, R. Guerrero, P. Bentley, R. N. Gunn, A. Hammers, D. A. Dickie, M. del C. Valdés Hernández,
281 J. M. Wardlaw, and D. Rueckert. Gan augmentation: Augmenting training data using generative
282 adversarial networks. *ArXiv*, abs/1810.10863, 2018.
- 283 A. Buades, B. Coll, and J. . Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society*
284 *Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65 vol. 2, June
285 2005. doi: 10.1109/CVPR.2005.38.
- 286 E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly
287 incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.
288 ISSN 0018-9448. doi: 10.1109/TIT.2005.862083.
- 289 J. Chen, J. Chen, H. Chao, and M. Yang. Image blind denoising with generative adversarial network based noise
290 modeling. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- 291 K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Bm3d image denoising with shape-adaptive principal
292 component analysis. *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations*
293 (*SPARS'09*), Apr 2009.
- 294 V. L. Dmitry Ulyanov, Andrea Vedaldi. Deep image prior. 1711.10925, 2017. URL <http://arxiv.org/abs/1711.10925>.
- 296 W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain
297 selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838–1857, July 2011.
298 ISSN 1057-7149. doi: 10.1109/TIP.2011.2108306.
- 299 D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289 – 1306, Apr 2006.
300 ISSN 0018-9448. doi: 10.1109/TIT.2006.871582.
- 301 P. Getreuer. Total Variation Inpainting using Split Bregman. *Image Processing On Line*, 2:147–157, 2012. doi:
302 10.5201/ipol.2012.g-tvi.
- 303 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio.
304 Generative adversarial nets. *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 2014.
- 305 S.-W. Huang, C.-T. Lin, S.-P. Chen, Y.-Y. Wu, P.-H. Hsu, and S.-H. Lai. Auggan: Cross domain adaptation with
306 gan-based data augmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer*
307 *Vision – ECCV 2018*, pages 731–744, Cham, 2018. Springer International Publishing. ISBN 978-3-030-
308 01240-3.
- 309 D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014. URL <http://arxiv.org/abs/1412.6980>. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International
310 Conference for Learning Representations, San Diego, 2015.
- 312 O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using
313 conditional adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*,
314 pages 8183–8192, June 2018. doi: 10.1109/CVPR.2018.00854.

- 315 E. Y. Lam and J. W. Goodman. Iterative statistical approach to blind image deconvolution. *J. Opt. Soc. Am. A*,
 316 17(7):1177–1184, Jul 2000. doi: 10.1364/JOSAA.17.001177. URL <http://josaa.osa.org/abstract.cfm?URI=josaa-17-7-1177>.
- 318 C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang,
 319 and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. pages
 320 105–114, 07 2017. doi: 10.1109/CVPR.2017.19.
- 321 Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International
 322 Conference on Computer Vision (ICCV)*, 2015.
- 323 X. Lu, W. Dong, P. Wang, G. Shi, and X. Xie. Convcsnet: A convolutional compressive sensing framework
 324 based on deep learning. *CoRR*, abs/1801.10342, 2018. URL <http://arxiv.org/abs/1801.10342>.
- 325 D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros. Context encoders: Feature learning by inpainting.
 326 In *CVPR*, 2016.
- 327 A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative
 328 adversarial networks. In *ICLR*, 2016.
- 329 W. Ruangsang and S. Aramvith. Efficient super-resolution algorithm using overlapping bicubic interpolation. In
 330 *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, pages 1–2, Oct 2017. doi: 10.1109/
 331 GCCE.2017.8229459.
- 332 J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert. A deep cascade of convolutional neural
 333 networks for mr image reconstruction. In M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap,
 334 and D. Shen, editors, *Information Processing in Medical Imaging*, pages 647–658, Cham, 2017. Springer
 335 International Publishing. ISBN 978-3-319-59050-9.
- 336 P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol. Stacked denoising autoencoders: Learning
 337 useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408,
 338 Dec. 2010. ISSN 1532-4435. URL <http://dl.acm.org/citation.cfm?id=1756006.1953039>.
- 339 D. Warde-Farley and Y. Bengio. Improving generative adversarial networks with denoising feature matching. In
 340 *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Workshop Track Proceedings*, 2017. URL <https://openreview.net/forum?id=S1X7nhsx1>.
- 342 L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. 2010.
- 343 Q. Yan and W. Wang. DCGANs for image super-resolution, denoising and deblurring. 2017.
- 344 R. Yan and L. Shao. Blind image blur estimation via deep learning. *IEEE Trans Image Process*, 25(4):1910–21,
 345 Apr 2016. doi: 10.1109/TIP.2016.2535273.
- 346 Y. Yang, J. Sun, H. Li, and Z. Xu. Deep admm-net for compressive sensing mri. In D. D. Lee,
 347 M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information
 348 Processing Systems 29*, pages 10–18. Curran Associates, Inc., 2016. URL <http://papers.nips.cc/paper/6406-deep-admm-net-for-compressive-sensing-mri.pdf>.
- 350 R. A. Yeh, C. Chen, T. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting
 351 with deep generative models. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR
 352 2017, Honolulu, HI, USA, July 21–26, 2017*, pages 6882–6890, 2017. doi: 10.1109/CVPR.2017.728. URL
 353 <https://doi.org/10.1109/CVPR.2017.728>.
- 354 J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention.
 355 *arXiv preprint arXiv:1801.07892*, 2018.
- 356 J. Zhang, D. Zhao, and W. Gao. Group-based sparse representation for image restoration. *IEEE Transactions on
 357 Image Processing*, 23(8):3336–3351, May 2014. ISSN 1057-7149. doi: 10.1109/TIP.2014.2323127.
- 358 J. J. Zhao, M. Mathieu, and Y. LeCun. Energy-based generative adversarial networks. In *ICLR*, 2017.