UNSUPERVISED ADVERSARIAL IMAGE RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

We address the problem of recovering an underlying signal from lossy and inaccurate measurements in an unsupervised fashion. Typically, we consider situations where there is no background knowledge on the structure of the unknown signal and where we do not have access to signal-measurement pairs, nor even unpaired signal data. We introduce a general framework, where a neural network is trained to recover plausible signals from the measurements in the data, by introducing an adversarial and a reconstruction loss. We evaluate our framework on different corruption instances, and show that our approach yields comparable results to model variants trained with stronger supervision.

1 INTRODUCTION

Many learning problems require acquiring information of the true state of some physical system from incomplete and noisy measurements. For example, weather forecasting relies on measurements acquired by different types of sensors. Data from these sensors are often incomplete and may be contaminated by external factors. For example, satellite sensors have to deal with the presence of clouds or with inherent errors relative to acquisition instruments, hindering the overall measurement process. In such cases, one needs to remove the effect of unwanted external factors and retrieve the true signal.

Recovering the signal and eliminating external factors from the measurements does not usually yield a unique solution, which means that multiple signals reconstructions could explain the measurements. For instance, if clouds are blocking the temperature acquisition, we could replace the missing temperature values with absurd ones, and this would explain the measurements just as accurately. One usually relies on some prior on the true signal in order to constrain the reconstruction to plausible solutions. A common approach is to hand-craft priors on the structure of the signal (Candès et al. (2005), Mota et al. (2017)). This approach is limited to model idealized situations in which the structure can be easily described, and which are rarely observed in the wild.

With the development of powerful generative models parameterized by neural networks (Goodfellow et al. (2014), Kingma & Welling (2013), Dinh et al. (2016)), a promising statistical approach to signal recovery is emerging, where priors are not handcrafted, but learned from large amounts of data. Despite exhibiting interesting results (Bora et al. (2017), Mardani et al. (2017), Ledig et al. (2016)), these methods all require some form of supervision, either measurement-signal pairs, or fully observed signal samples. For many practical problems, *e.g.* scientific data acquisition mentioned above, obtaining these samples is extremely expensive and/or impractical, which makes these approaches unrealistic for such situations.

In this work, we propose a general framework allowing to recover signals from noisy and lossy data, using only a dataset of lossy measurements. We compare our approach with baselines and variants of our model that have access to fully observed signal data, or to measurement-signal pairs.

2 PRELIMINARIES

Notations. We use capital letters (*e.g.* X) for random variables, and lower-case letters (*e.g.* x) for their values. $\mathrm{p}_X(x)$ denotes the distribution (or the density in the appropriate context) of X evaluated at x.

2.1 PROBLEM SETTING.

We suppose that a measurement $Y \sim \mathrm{p}_Y$ is generated jointly from a signal $X \sim \mathrm{p}_X$ and from additional *known* external factors $\Theta \sim \mathrm{p}_\Theta$ combined with a measurement operator $F : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^m$. Additional sources of uncertainty due to *unknown* factors, or due to the fact that F is imperfect can be modeled with an additive Gaussian noise $\mathcal{E} \sim \mathcal{N}(0, \sigma^2 I)$:

$$Y = F(X; \Theta) + \mathcal{E} \tag{1}$$

We assume F to be differentiable *w.r.t.* its first argument X, and p_Θ to be easy to sample from, and Θ and X to be independent. If $F(X; \Theta)$ adequately explains the measurements, we can choose $\mathcal{E}$ such that σ^2 is close to zero. For example, Θ could correspond to the position or the size of the clouds hindering the measurements, and $\mathcal{E}$ to instrumental noise or other factors that are not modelled with F. We wish to find a general method that recovers the signal X from the lossy, inaccurate measurements Y.

2.2 APPROACH

The posterior distribution $\mathrm{p}_{X|Y}(\cdot|y)$ describes the current state of knowledge we have about the signal when measurement y is observed. *Maximum a posteriori* (MAP) estimation consists in retrieving the most probable signal $\hat{x}$ under this distribution:

$$\hat{x} = \arg\max_x \log \mathrm{p}_{X|Y}(x|y) \tag{2}$$

In order to recover signal x from measurement y, we want to learn a function $G : \mathcal{Y} \to \mathcal{X}$ such that each measurement y is associated with its MAP estimate $\hat{x}$. Therefore, plugging estimate $G(y)$ in place of the true signal value$x$, it is natural to maximize the following average log-posterior:

$$\arg\max_G \mathbb{E}_{\mathrm{p}_Y} \log \mathrm{p}_{X|Y}(G(y)|y) \tag{3}$$

Applying Bayes' theorem to the posterior and ignoring the normalizing constant, objective 3 is equivalent to:

$$\arg\max_G \mathbb{E}_{\mathrm{p}_Y} \left\{ \log \mathrm{p}_{Y|X}(y|G(y)) + \log \mathrm{p}_X(G(y)) \right\} \tag{4}$$

where $\mathrm{p}_{Y|X}$ is the likelihood function and p_X is the prior. Therefore, the problem amounts at finding a reconstruction function $G(y)$ that is: (1) likely to have generated the data, *i.e.* yielding high probability under the likelihood $\mathrm{p}_{Y|X}$ and (2) is plausible, *i.e.* yielding high probability under the prior p_X. Unfortunately, these two distributions are difficult to evaluate: computing the likelihood requires computing an intractable integral, and the prior distribution is unknown overall. In the following sections, we will see how to handle these two terms.

3 METHOD

We wish to learn a reconstruction network G that takes as input corrupted measurements and produces reconstructions that yield high probability for the likelihood and the prior. The model will be trained using a combination of two losses, one for each requirement, as is shown in Figure 1.

3.1 HANDLING THE LIKELIHOOD TERM

In the general case, evaluating the likelihood $\mathrm{p}_{Y|X}(y|x)$ of a signal with value x given measurement y requires marginalizing on the unobserved noise variable Θ: $\mathrm{p}_{Y|X}(y|x) = \mathbb{E}_{\mathrm{p}_\Theta}\left\{\mathrm{p}_{Y|X,\Theta}(y|x,\theta)\right\}$, which involves computing an intractable integral. Many probabilistic approaches for image denoising design the measurement model in order to obtain a

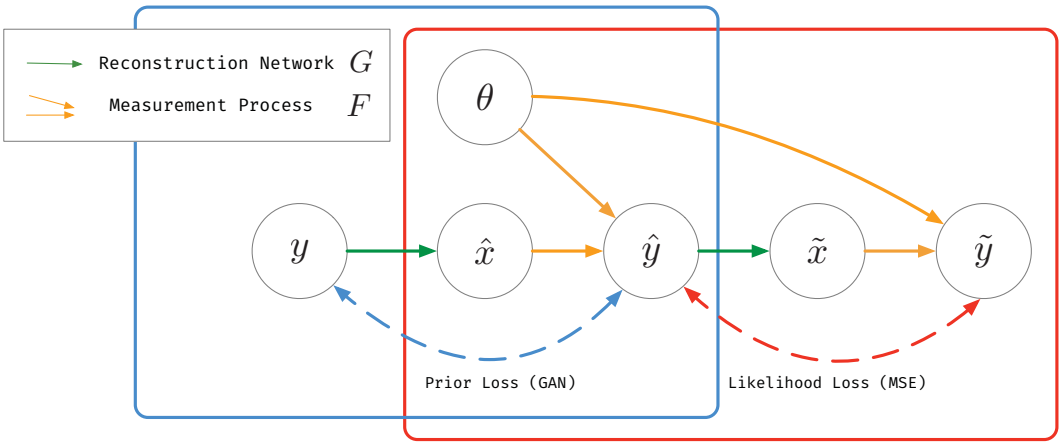

Figure 1: Schematic representation of the proposed approach. We wish to train G to recover a plausible signal from lossy measurements. As is shown in Section 2.2, this requires the reconstructions $\hat{x} := G(y)$ to have high probability under the likelihood (*Red frame*) and the prior (*Blue frame*). For simplicity, variable $\mathcal{E}$ has been omitted. *Blue frame:* we sample a measurement y from the data, produce a reconstruction $\hat{x}$, and sample a perturbation parameter θ. We enforce the simulated measurement $\hat{y} := F(\hat{x}; \theta)$ to be similar to measurements in the data using an adversarial penalty. Intuitively, this requires the network to *remove* the corruption. *Red frame:* to enforce G to produce reconstructions with high likelihood, it is not possible to add a penalty to constrain the mean square error (MSE) between y and $\hat{y}$ to be small. This is because the underlying perturbation that caused y is unknown, and may be different from θ. However, if $\hat{y}$ is similar to the measurements in the data, we can use it as a proxy for a measurement of the data where its associated θ is known. We can now constrain $\|\hat{y} - F(G(\hat{y}); \theta)\|_2^2$ to be small.

known analytic form (Boyat & Joshi (2015), Alkinani & El-Sakka (2017)). However, we wish to attack the general problem and therefore choose not to make this assumption.

We will show that thanks to the independence assumption between Θ and X, the log-likelihood can be transformed into an amenable form for us. Making use of this independence assumption, we can decompose the log-likelihood as:

$$\begin{aligned} \log \mathrm{p}_{Y|X}(y|x) &= \log \mathrm{p}_{Y,\Theta|X}(y, \theta|x) - \log \mathrm{p}_{\Theta|X,Y}(\theta|x, y) \\ &\overset{X \perp\!\!\!\perp \Theta}{=} \log \mathrm{p}_{Y|X,\Theta}(y|x, \theta) + \log \mathrm{p}_{\Theta}(\theta) - \log \mathrm{p}_{\Theta|Y}(\theta|y) \end{aligned} \tag{5}$$

Applying the expectation *w.r.t* to $\Theta|Y$ on both sides, we obtain

$$\log \mathrm{p}_{Y|X}(y|x) = \mathbb{E}_{\mathrm{p}_{\Theta|Y}} \left\{ \log \mathrm{p}_{Y|X,\Theta}(y|x, \theta) + \log \mathrm{p}_{\Theta}(\theta) - \log \mathrm{p}_{\Theta|Y}(\theta|y) \right\} \tag{6}$$

Since the last two terms do not depend on the signal, this means that in equation 4, where x is substituted with $G(y)$, this term has no impact and can be ignored. Thus, handling the likelihood term in equation 4 is equivalent to handling $\mathbb{E}_{\mathrm{p}_Y \mathrm{p}_{\Theta|Y}} \left\{ \log \mathrm{p}_{Y|X,\Theta}(y|G(y), \theta) \right\}$. This form is convenient since $\log \mathrm{p}_{Y|X,\Theta}(y|x, \theta)$ yields a simple analytic expression:

$$\log p(y|x, \theta) = -\frac{1}{2\sigma^2} \|y - F(x; \theta)\|_2^2 + c \tag{7}$$

where c is a constant. A natural penalty to enforce high likelihood is thus:

$$\begin{aligned} \mathcal{L}^{\text{likeli}}(G) &:= \mathbb{E}_{\mathrm{p}_\Theta \mathrm{p}_{Y|\Theta}} \|y - F(G(y); \theta)\|_2^2 \\ &= \mathbb{E}_{\mathrm{p}_\Theta \mathrm{p}_X \mathrm{p}_{Y|X,\Theta}} \|y - F(G(y); \theta)\|_2^2 \end{aligned} \tag{8}$$

We have taken the expectation *w.r.t* $\mathrm{p}_\Theta \mathrm{p}_{Y|\Theta}$ instead of $\mathrm{p}_Y \mathrm{p}_{\Theta|Y}$, which from a theoretical viewpoint is equivalent. In practice, this allows us to sample from p_Θ, and then sample from $\mathrm{p}_{Y|\Theta}$. If the signal distribution p_X were known, given that $\mathrm{p}_{Y|\Theta}(y|\theta) = \mathbb{E}_{\mathrm{p}_X}\big\{\mathrm{p}_{Y|X,\Theta}(y|x,\theta)\big\}$ the expectation in equation 8 could be computed. However, p_X is unknown. In the following sections, we will see how we work around this problem. This process is associated to the red frame in Figure 1.

3.2 Handling the Prior term

How can we generate reconstructed signals that are probable under the prior if the latter is unknown? If we had access to fully-observed signal data, it could be possible to enforce our reconstruction network to produce signals that are indistinguishable from the ones in the data by leveraging adversarial training, similarly to what is done in the CycleGAN model (Zhu et al. (2017)). However, as stated above, we do not have access to the true signal.

We will build on an idea introduced in the AmbientGAN model by Bora et al. (2018). AmbientGAN aims at learning an *unconditional* generative model of the true signal distribution, when only lossy measurements of the signal are available. They train a generator to produce signal samples from a latent code, which when corrupted are indistinguishable from the measurements in the data *w.r.t.* a discriminating network. Intuitively, this requires the network to produce signal that are uncorrupted. Bora et al. (2018) shows that given a fixed noise distribution p_Θ, if the distribution of measurements uniquely determines the distribution of signals, Nash Equilibrium for the adversarial game must imply that the generator's induced distribution matches the signal's true distribution (for more details on the measurements for which there exists provable guarantees, see Section 4.2, or Bora et al. (2018)). Note that this is different from our problem: we want to reconstruct corrupted signals and not to learn a distribution of these signals. In AmbientGAN, the generator only takes as input a latent code, and reconstruction from a specific measurement is not possible.

Taking inspiration from this approach, we add an adversarial penalty to enforce our *conditioned* generator G to produce reconstructions, such that their associated measurements are indistinguishable from the measurements in the data *w.r.t.* a discriminator D. We introduce the following penalty:

$$\mathcal{L}^{\text{prior}}(G) := \max_D \; \mathbb{E}_{Y\sim \mathrm{p}_Y, \hat{Y}\sim \mathrm{p}_Y^G}\big\{\log D(y) + \log\big(1 - D(\hat{y})\big)\big\} \tag{9}$$

where p_Y^G corresponds to the distribution induced by G's corrupted outputs, *i.e.* $\mathrm{p}_Y^G(y) := \mathbb{E}_{\mathrm{p}_\Theta \mathrm{p}_X^G}\big\{\mathrm{p}(y|x,\theta)\big\}$ and p_X^G denotes the marginal distribution induced by G's outputs: $\mathrm{p}_X^G(x) := \mathbb{E}_{\mathrm{p}_Y} \mathrm{p}_{X|Y}^G(x|y) = \mathbb{E}_{\mathrm{p}_Y} \delta\big(x - G(y)\big)$[1]. This penalty enforces the marginal p_X^G to be close to the true prior distribution p_X, and thus enforces G to map its input measurements onto p_X. This process is associated to the blue frame in Figure 1.

3.3 In practice

In Section 3.1, we have shown that it possible to maximize the average log-likelihood, given that we can sample from the unknown prior distribution p_X. In Section 3.2, we have shown how it is possible to enforce the generator to produce signals from p_X without ever having access to uncorrupted samples. We can then use the distribution induced by the generator's output p_X^G as a proxy for p_X to compute an approximate value of the expectation in equation 8. Taking the expectation *w.r.t.* the distribution p_X^G induced by the generator instead of p_X gives us the approximate penalty term:

$$\mathcal{L}^{\widehat{\text{likeli}}}(G) := \mathbb{E}_{\mathrm{p}_\Theta \mathrm{p}_X, \hat{Y}\sim \mathrm{p}_{Y|X,\Theta}^G} \|\hat{y} - F(G(\hat{y});\theta)\|_2^2 \tag{10}$$

The full objective is a linear combination of both penalties:

$$\arg\min_G \; \mathcal{L}^{\text{prior}}(G) + \lambda \cdot \mathcal{L}^{\widehat{\text{likeli}}}(G) \tag{11}$$

[1] $\delta(x)$ is the Dirac delta function, which is equal to zero everywhere except in x.

Under appropriate conditions introduced in Section 3.2, if the Nash Equilibrium for the adversarial game is reached, $\mathrm{p}_X^G = \mathrm{p}_X$ is recovered, which then implies that the approximate likelihood penalty $\widehat{\mathcal{L}^{\text{likeli}}}(G)$ reduces to the likelihood penalty $\mathcal{L}^{\text{likeli}}(G)$.

In practice, we use a Monte-Carlo estimation to approximate the expectations for both terms in 11, as described in Figure 1 and Algorithm 1.

Algorithm 1 Training Procedure.

Require: Initialize parameters of the the generator G and the discriminator D.

while (G, D) not converged **do**

Sample $\{y_i\}_{1\leq i\leq n}$ from data distribution p_Y

Sample $\{\theta_i\}_{1\leq i\leq n}$ from P_Θ

Sample $\{\varepsilon_i\}_{1\leq i\leq n}$ from $P_\mathcal{E}$

Set $\hat{y}_i$ to $F(G(y_i), \theta_i) + \varepsilon_i$ for $1 \leq i \leq n$

Update D by ascending:

$$\frac{1}{n}\sum_{i=1}^{n} \log D(y_i) + \log(1 - D(\hat{y}_i))$$

Update G by descending:

$$\frac{1}{n}\sum_{i=1}^{n} \lambda \cdot \|\hat{y}_i - F(G(\hat{y}_i); \theta_i)\|_2^2 - \log D(\hat{y}_i)$$

end while

4 EXPERIMENTS

4.1 ARCHITECTURE DETAILS

Network. For all are experiments, we use the Resnet architecture He et al. (2016) for our reconstruction network G. For the measurement instance *Patch Band* (see Section 4.2), we use $G(y) = y + \text{Resnet}(y)$ and use small initial parameters in order to initialize our network to the identity function. This allows our model to learn the distribution of the *difference* of the signal and measurements, which is simpler for the model when the measurements are similar to the associated signal. For the discriminator D, we used the Patch-GAN architecture (Zhu et al. (2017); Isola et al. (2016)).

The hyper-parameters of the models, *e.g.* the number of layers or the dropout rate are chosen using the validation set. Due to the large number of combinations of hyper-parameters, we made a *random search* (Bergstra & Bengio (2012)) and selected the model with the lowest mean square error between the reconstructions $\hat{x}$ and the image x.

Training. We use the Adam solver (Kingma & Ba (2014)) with a batch size of 32. All networks were trained from scratch with a cross validated learning rate. We linearly decay the learning rate between epochs 20 and 100.

We evaluate our approach on the CelebA dataset (Liu et al. (2015)), which consists in 200 000 celebrity photos [2]. We use the same preprocessing of the data as Bora et al. (2018): the images are cropped and resized to 64×64. The images have all been corrupted once, *i.e.* there is no two measurements in the data generated with the same uncorrupted image.

4.2 CORRUPTIONS

As for the corruption function, we used the following measurement model :

[2] The dataset is available at `http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html`.

Remove-Pixel In this measurement process, 95% of pixels are randomly selected and set to 0.

Remove-Pixel-Channel In this measurement process 95% of channel values are randomly selected and set to 0.

Convolve-Noise The image is convoluted with a Gaussian kernel of size k, and a Gaussian noise of variance $\sigma = 0.15$.

Patch-Band For this function, a horizontal band of height h is randomly set to zero. It can be similar to cloud weather data completion.

Influence of the corruption function can be seen in the figures in section 4. For measurement models *Remove-Pixel*, *Convolve-Noise* and *Remove-Pixel-Channel*, Bora et al. (2018) show that the distribution of measurements induces a unique distribution of signals.

4.3 Baselines

Conditional AmbientGan. An unconditional generator is trained using the AmbientGan framework (Bora et al. (2018)) to produce samples from the uncorrupted signal distribution, for each type of corruption (see Section 4.2). We use the same architecture and hyperparameters as those provided by the authors. For each measurement y, we look for the latent code z that produced y, and selects the reconstruction $\hat{x}$ that solves $\hat{x} = \arg\min_z \|y - G(z)\|_2^2$, as is proposed in Bora et al. (2017). This approach is the most similar to ours, and does not require having access to samples from the signal distribution. Because this approach is sensitive to the initial latent code, we reiterate this approach three times and select the best resulting image.

Unpaired Supervision This is a variant of our model that has access to the fully-observed images. This model is identical to ours, apart from the fact that instead of discriminating between measurements y and generated measurements $\hat{y}$, we discriminate between fully-observed images x and generated images $\hat{x}$.

Paired supervision This is a variant of our model that has access to pairs of measurements and fully-observed images (y, x). We discriminate between fully-observed images x and generated images $\hat{x}$, and regress to the associated fully-observed image x, like in the Pix2Pix model (Isola et al. (2016)).

Other. We also compare our model to different baselines, including Deep Image Prior (Ulyanov et al. (2017)), Damelin & Hoang (2018), and *Total Variation Denoising*[3]. These results are provided in the appendix.

4.4 Quantitative Results

We now compare our model with those presented in the previous section. Each model is evaluated with a mean square error metric between the reconstructed $\hat{x}$ and the x used to generate the input y. The hyper-parameters are tuned on the validation set. In Table 1, we report the performance on the test set, which is a random subset of the CelebA dataset of size 40000. Because the Conditional AmbientGan model is computationally expensive, we only report the MSE on 40 randomly chosen samples of the test set.

Table 1: : Average MSE on the test set. Each column corresponds to a different corruption process.

	Remove-Pixel	Remove-Pixel-Channel	Patch-Band	Convolve-Noise
Conditional AmbientGan	0.292	0.2829	0.1421	0.0814
Ours	0.0414	0.0409	0.0165	0.0088
Unpaired Supervision	**0.037**	**0.0336**	0.034	0.0103
Paired Supervision	0.0383	0.0401	**0.0147**	**0.0084**

[3] http://scikit-image.org/docs/0.12.x/auto_examples/filters/plot_denoise.html

Quantitatively, our model performs well. Except for the Conditional Ambiant GAN, all the methods are quite similar in MSE. This means that our unsupervised model reaches performance very close - sometimes better - than the two supervised models.

We remark that even if we do not have access to aligned pair, computing the prior loss directly on the uncorrupted signal yield results comparable, and sometimes better than the paired supervision model. This empirically suggests that the likelihood loss is sufficient to efficiently condition the reconstruction on the input signal.

4.5 Qualitative Results

Besides MSE, we need to analyze the reconstructed samples qualitatively. Figure 2 shows the reconstruction obtained by the different model. Each row corresponds to a kind of measurement, and each column to a different model. As we can see, the measurement function we used, induce a large loss of information, and is difficult to reconstruct, even for a human.

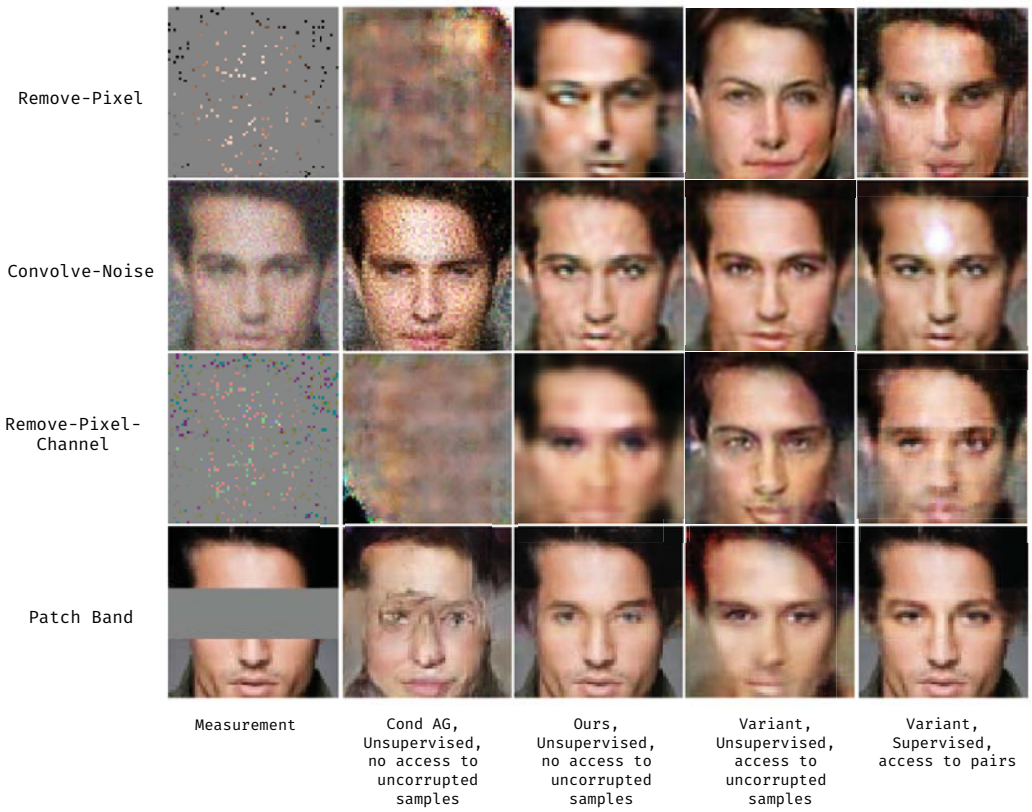

Figure 2: Reconstructions for each of the models. Each row corresponds to a specific corruptions, and each columns to a specific model.

The visual reconstruction is coherent with the quantitative results. In figure 3 and 4 we show randomly sampled reconstruction from our model.

5 Related Work

Up to our knowledge, there is no other Deep Learning approach attempting to solve the unsupervised signal reconstruction problem. However, some of the ideas developed here are close or even inspired as mentioned in the text to recent existing work.

The idea of cycle consistency has been proposed or used in several contexts. (Zhu et al. (2017), Lample et al. (2017), Almahairi et al. (2018)) for example all exploit this constraint in order to learn from unpaired data sets. To enforce high likelihood, we recover a penalty loss that is close to

Figure 3: Random samples from our model, trained to reconstruct the images from Patch-Band measurements.

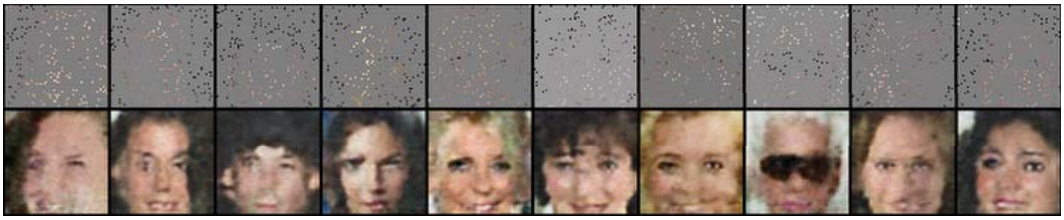

Figure 4: Random samples from our model, trained to reconstruct the images from Remove-Pixel measurements.

the Cycle consistency loss. Moreover, they too use, as we did, adversarial training to constrain the marginal distribution induced by the generator.

In the context of image super resolution, (Ledig et al. (2016), Sønderby et al. (2016), Mardani et al. (2017)) attempt to retrieve map estimates of the super resolution image conditioned on an input image. They too use a generative model of the signal trained in an adversarial fashion using samples from signal distribution to constrain their reconstructions. Their approach is fully supervised.

Other works attempt to solve ill-posed inverse problems using generative models (Bora et al. (2017), Ulyanov et al. (2017), Asim et al. (2018), Tripathi et al. (2018), Van Veen et al. (2018)). The general approach in all these contributions consists in inverting the generative model by finding a latent code close to the measurement by directly minimizing the mean square error between the corrupted reconstruction and the measurement. This requires solving an optimization problem for each image, which takes several minutes (Ulyanov et al. (2017)) on GPU, and requires random restarts to avoid falling a bad local minima. Again, the setting is fully supervised.

Finally Lehtinen et al. (2018) propose a method for denoising images without direct supervision. They train a network to regress a corrupted image using the same image with a different corruption value as input. Assuming the corruption has zero-mean, their network learns to remove the corruption which is present in the input image. This setting implies that they have access to the original (uncorrupted) image for generating different noisy versions of the same image, which is not our case. Our setting is then more complex and probably more realistic.

6 Conclusion

We have proposed a general formulation to recover a signal from lossy measurements using a neural network, without having access to uncorrupted signal data. From the *maximum a posteriori* approach, we derive two penalties for our neural network. By minimizing a linear combination of the previous penalties, we show that our approach yields results that are competitive with other models that have access to higher forms of supervision.

For future work, we plan to apply our framework to different corruption processes, and test if our model holds in the wild, specifically for retrieving uncorrupted scientific data. Lastly, it would be interesting to approximate the true posterior by making our model stochastic.

REFERENCES

Monagi H. Alkinani and Mahmoud R. El-Sakka. Patch-based models and algorithms for image denoising: a comparative review between patch-based images denoising methods for additive noise reduction. *EURASIP Journal on Image and Video Processing*, 2017(1):58, Aug 2017. ISSN 1687-5281. doi: 10.1186/s13640-017-0203-4. URL https://doi.org/10.1186/s13640-017-0203-4.

Amjad Almahairi, Sai Rajeswar, Alessandro Sordoni, Philip Bachman, and Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. *arXiv preprint arXiv:1802.10151*, 2018.

Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Solving bilinear inverse problems using deep generative priors. *CoRR*, abs/1802.04073, 2018. URL http://arxiv.org/abs/1802.04073.

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed Sensing using Generative Models. *arXiv:1703.03208 [cs, math, stat]*, March 2017. URL http://arxiv.org/abs/1703.03208. arXiv: 1703.03208.

Ashish Bora, Eric Price, and Alexandros G. Dimakis. AmbientGAN: Generative models from lossy measurements. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Hy7fDog0b.

Ajay Kumar Boyat and Brijendra Kumar Joshi. A review paper: Noise models in digital image processing. *CoRR*, abs/1505.03489, 2015.

Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2005. doi: 10.1002/cpa.20124. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20124.

SB Damelin and NS Hoang. On surface completion and image inpainting by biharmonic functions: Numerical aspects. *International Journal of Mathematics and Mathematical Sciences*, 2018, 2018.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. *CoRR*, abs/1605.08803, 2016. URL http://arxiv.org/abs/1605.08803.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pp. 2672–2680, Cambridge, MA, USA, 2014. MIT Press. URL http://dl.acm.org/citation.cfm?id=2969033.2969125.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pp. 770–778, 2016. URL https://doi.org/10.1109/CVPR.2016.90.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. URL http://arxiv.org/abs/1611.07004.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2013. URL http://arxiv.org/abs/1312.6114.

Guillaume Lample, Ludovic Denoyer, and Marc'Aurelio Ranzato. Unsupervised machine translation using monolingual corpora only. *CoRR*, abs/1711.00043, 2017. URL http://arxiv.org/abs/1711.00043.

Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, abs/1609.04802, 2016. URL http://arxiv.org/abs/1609.04802.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning Image Restoration without Clean Data. *arXiv:1803.04189 [cs, stat]*, March 2018. URL http://arxiv.org/abs/1803.04189. arXiv: 1803.04189.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

Morteza Mardani, Enhao Gong, Joseph Y. Cheng, Shreyas Vasanawala, Greg Zaharchuk, Marcus T. Alley, Neil Thakur, Song Han, William J. Dally, John M. Pauly, and Lei Xing. Deep generative adversarial networks for compressed sensing automates MRI. *CoRR*, abs/1706.00051, 2017. URL http://arxiv.org/abs/1706.00051.

João FC Mota, Nikos Deligiannis, and Miguel RD Rodrigues. Compressed sensing with prior information: Strategies, geometry, and bounds. *IEEE Transactions on Information Theory*, 63(7): 4472–4496, 2017.

Casper Kaae Sønderby, Jose Caballero, Lucas Theis, Wenzhe Shi, and Ferenc Huszár. Amortised MAP Inference for Image Super-resolution. *arXiv:1610.04490 [cs, stat]*, October 2016. URL http://arxiv.org/abs/1610.04490. arXiv: 1610.04490.

Subarna Tripathi, Zachary C. Lipton, and Truong Q. Nguyen. Correction by projection: Denoising images with generative adversarial networks. *CoRR*, abs/1803.04477, 2018. URL http://arxiv.org/abs/1803.04477.

Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. *CoRR*, abs/1711.10925, 2017. URL http://arxiv.org/abs/1711.10925.

David Van Veen, Ajil Jalal, Eric Price, Sriram Vishwanath, and Alexandros G. Dimakis. Compressed Sensing with Deep Image Prior and Learned Regularization. *arXiv:1806.06438 [cs, math, stat]*, June 2018. URL http://arxiv.org/abs/1806.06438. arXiv: 1806.06438.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *arXiv:1703.10593 [cs]*, March 2017. URL http://arxiv.org/abs/1703.10593. arXiv: 1703.10593.

Figure 5: Baseline comparison for the remove-pixel measurement

Figure 6: Baseline comparison for the remove-pixel-channel measurement

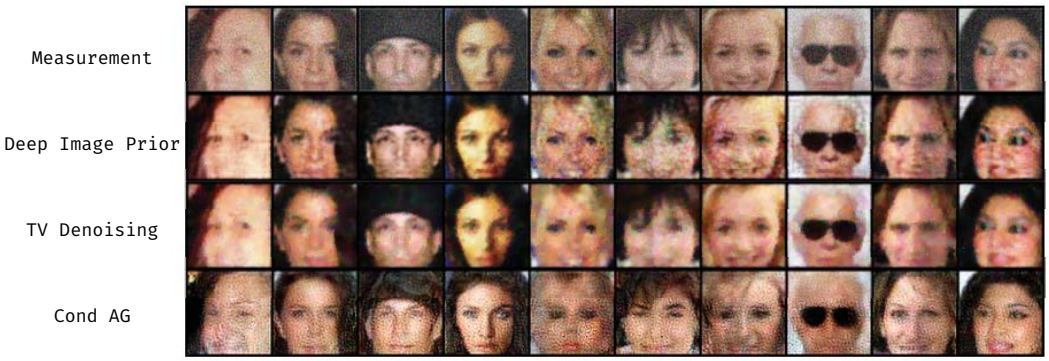

Figure 7: Baseline comparison for the convolve-noise measurement

Figure 8: Baseline comparison for the Patch-Band measurement

Figure 9: Random samples from our model, trained to reconstruct the images from Convolve-Noise measurements with a noise variance of 0.15 .

Figure 10: Random samples from our model, trained to reconstruct the images from Convolve-Noise measurements with a noise variance of 0.5 .

Figure 11: Random samples from our model, trained to reconstruct the images from keep-patch measurements.

Figure 12: Random samples from our model, trained to reconstruct the images from remove-pixel measurements with $p = 0.80$.

Figure 13: Random samples from our model, trained to reconstruct the images from remove-pixel measurements with $p = 0.95$.

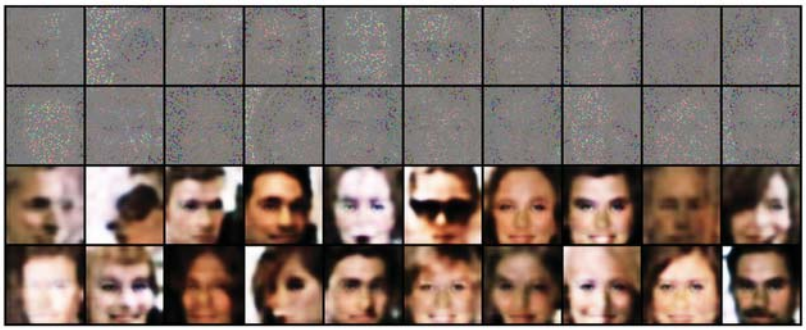

Figure 14: Random samples from our model, trained to reconstruct the images from remove-pixel-channel measurements with $p = 0.95$.