ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

Pixel-wise Conditioning of Generative Adversarial Networks

Cyprien Ruffino[1], Romain Hérault[1], Eric Laloy[2], Gilles Gasso[1] *

1- Normandie Univ, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS
76 000 Rouen, France

2- Belgian Nuclear Research, Institute Environment, Health and Safety,
Boeretang 200 - BE-2400 Mol, Belgium

Abstract. Generative Adversarial Networks (GANs) have proven successful for unsupervised image generation. Several works extended GANs to image inpainting by conditioning the generation with parts of the image one wants to reconstruct. However, these methods have limitations in settings where only a small subset of the image pixels is known beforehand. In this paper, we study the effectiveness of conditioning GANs by adding an explicit regularization term to enforce pixel-wise conditions when very few pixel values are provided. In addition, we also investigate the influence of this regularization term on the quality of the generated images and the satisfaction of the conditions. Conducted experiments on MNIST and FashionMNIST show evidence that this regularization term allows for controlling the trade-off between quality of the generated images and constraint satisfaction.

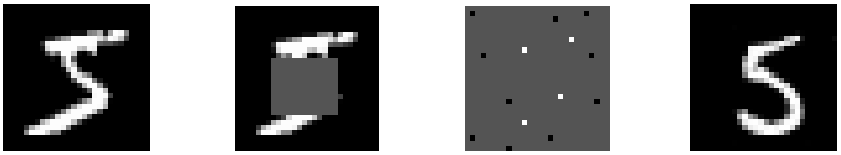
1 Introduction

In this work we consider an extreme setting of inpainting task: we assume that only a few pixels, less than a percent of the considered image size, are known and that these pixels are randomly scattered across the image (see Fig.1c). This raises the challenge of how to take advantage of this scarce and unstructured a priori information to generate high quality images. Besides methodological novelty, a method that can tackle this problem would find applications to GAN-based geostatistical simulation and inversion in the geosciences [1].

More specifically, this paper proposes an extension of the Conditional Generative Adversarial Network (CGAN) [2] framework to learn the distribution of the training images given the constraints (the known pixels). To make the generated images honoring the prescribed pixel values, we use a regularization term measuring the distance between the real constraints and their generated counterparts. Thereon we derive a learning scheme and analyze the influence of the used regularization term on both the quality of the generated images and the fulfillment of the constraints. By experimenting with a wide range of values for the additional hyper-parameter introduced by the regularization term, we show for the MNIST [3] and FashionMNIST [4] datasets that our approach is effective and allows for controlling the trade-off between the quality of the generated samples and the satisfaction of the constraints.

*This research was supported by the CNRS PEPS I3A REGGAN project and the ANR-16-CE23-0006 grant *Deep in France*

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.



(a) Original Image (b) Regular Inpainting (c) Scatter Inpainting (d) Generated Image

Fig. 1: Difference between regular inpainting (b) and the problem undertaken in this work (c). The image obtained with our framework is shown in (d).

2 Related works

Generative Adversarial Networks [5] basically consist of an algorithm for training generative models in an unsupervised way. It relies on a game between a generator, G, and a discriminator network, D, in which G learns to produce new data with similar spatial characteristics/patterns as in the true data while D learns to distinguish real examples from generated ones. Training GANs is equivalent to finding a Nash equilibrium to the following mini-max game:

$$\min_G \max_D L(D,G) = \mathop{\mathbb{E}}_{x\sim P_r}\Big[\log(D(x))\Big] + \mathop{\mathbb{E}}_{z\sim P_z}\Big[\log(1-D(G(z)))\Big] \qquad (1)$$

where P_z is a known distribution, usually normal or uniform, in which latent variables are drawn, and P_r is the distribution of the real samples.

Yeh et al. [6] introduced an inpainting method which consists of taking a pre-trained generator and exploring its latent space $\mathcal{Z}$ via gradient descent, to find a latent vector, z, which induces an image close to the altered one while its quality remains close to the real samples. This method was applied by Mosser et al. [7] for 3D image completion with few constraints. However, the location of the constraints in their approach was fixed, instead of randomly scattered.

Some other approaches rely on Conditional Generative Adversarial Networks (CGAN) [2]. This is a variant of GANs in which additional information, c, is given to both the generator and the discriminator as an input (see Fig.2a). The optimization problem becomes:

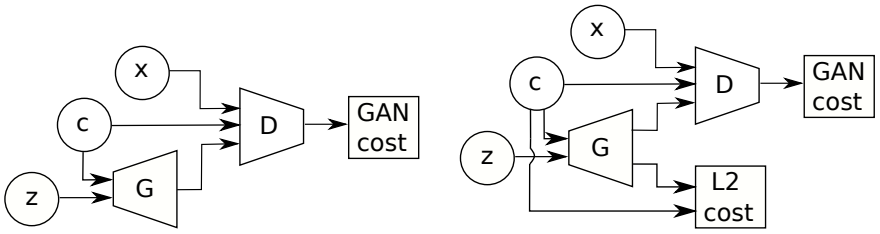
$$\min_G \max_D L(D,G) = \mathop{\mathbb{E}}_{\substack{x\sim P_r \\ \tilde{c}\sim P_{c|x}}}\Big[\log(D(x,\tilde{c}))\Big] + \mathop{\mathbb{E}}_{\substack{z\sim P_z \\ c\sim P_c}}\Big[\log(1-D(G(z,c),c))\Big] \qquad (2)$$

In it seminal version [2], CGANs are used for class-conditioned image generation by giving the labels of the images to the networks. However, several kind of conditioning data can be used even a full image to do image-to-image translation [8] or image inpainting [9, 10].

3 Proposed approach

In this work, we retain the CGAN approach and add a reconstruction loss term to further enforce the prescribed pixel values (usually less than a percent of the

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.



(a) CGAN (b) Our approach

Fig. 2: Different GAN Setups

image). With this setup, the generator can be used to generate images from constraints unseen during the training.

Given a learning set of images $X \in [-1,1]^{P\times P}$ drawn from an unknown distribution P_r and a sparse matrix $C \in [-1,1]^{P\times P}$ as the given constrained pixels, the problem we focused on consists in finding a generative model G with input $z \sim P_z$, a random vector sampled from a known distribution, and constrained pixel values $\tilde{C} \in [-1,1]^{P\times P}$ that could generate an image satisfying the constraints while likely following the distribution P_r. Enforcing the constraints in the CGAN framework leads to the following problem:

$$\min_G \max_D L(D,G) = \mathop{\mathbb{E}}_{\substack{X\sim P_r \\ \tilde{C}\sim P_{C|X}}} \Big[\log(D(X,\tilde{C}))\Big] + \mathop{\mathbb{E}}_{\substack{z\sim P_z \\ C\sim P_C}} \Big[\log(1-D(G(z,C),C))\Big] \quad (3)$$
$$\text{s.c. } C = M(C)\odot G(z,C))$$

where $\odot$ is the Hadamard (or point-wise) product and $M(C)$ is a corresponding masking matrix. $M(C)$ is a sparse matrix with entries equal to one at constrained pixels location. As the equality constraint in 3 is hard to enforce during training, we rather investigate a relaxed version of the problem. Indeed, we minimize the L_2 norm between the constrained pixels and the generated values (see Fig.2b). The objective function, with $\lambda \geq 0$ a regularization parameter, becomes:

$$\begin{aligned} L(D,G) \quad = \quad & \mathop{\mathbb{E}}_{\substack{X\sim P_r \\ \tilde{C}\sim P_{C|X}}} \Big[\log(D(X,\tilde{C}))\Big] \\ & + \mathop{\mathbb{E}}_{\substack{z\sim P_z \\ C\sim P_C}} \Big[\log(1-D(G(z,C),C)) + \lambda \left\|C - M(C)\odot G(z,C)\right\|_2^2\Big]. \end{aligned} \quad (4)$$

4 Experiments

We experiment on the MNIST [3] and FashionMNIST [4] datasets, which consist of images of size 28×28px. We split the official training set into a new training set (90%) and a validation set (10%). The official test set remains our test set. A fifth of each so defined set is used to generated the matrix of constraints C

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

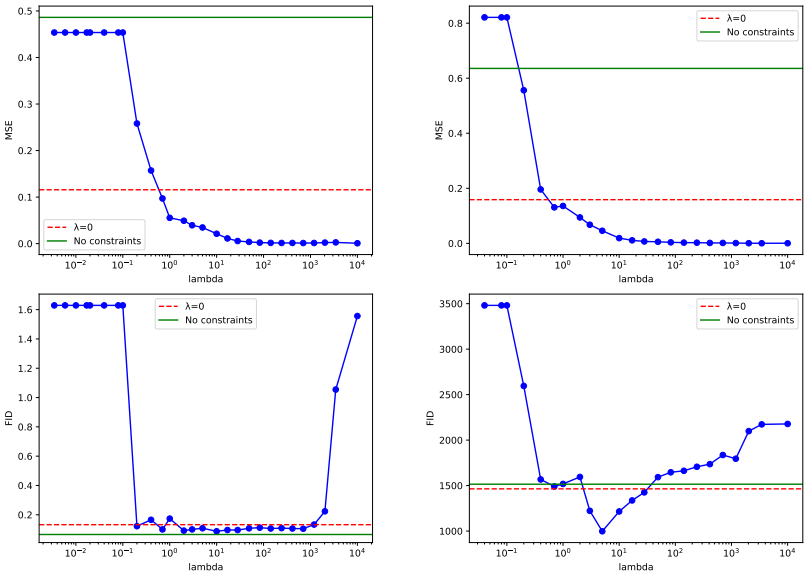


Fig. 3: MSE (top) and FID (bottom) w.r.t. the regularization parameter λ; Dataset MNIST (left), Fashion MNIST (right).

by randomly selecting 0.5% of the pixels. These images are then removed from the training sets, to avoid correlation between real example presented to the discriminator and constrained maps given to the generator.

A discriminator such as presented in DCGAN [11] has been chosen with only two convolutional layers of 64 and 128 filters, Leaky ReLU activations and batch normalization [12]. For the generator we retain the DCGAN architecture with a fully-connected layer and two transposed convolutional layers of 128 and 64 filters with ReLU activations and batch normalization. An example of a generated image with the corresponding constraints can be seen in figures 4c and 4d.

(a) Original Image (b) Constraints (c) Generated Image (d) Satisfied Consts.

Fig. 4: Generation of a sample during training. We first sample an image from a training set (a) and we sample the constraints from it. Then our GAN generates a sample (c). The constraints with squared error smaller than $\epsilon = 0.1$ are deemed satisfied and shown by green pixels in (d) while the red pixels are unsatisfied.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

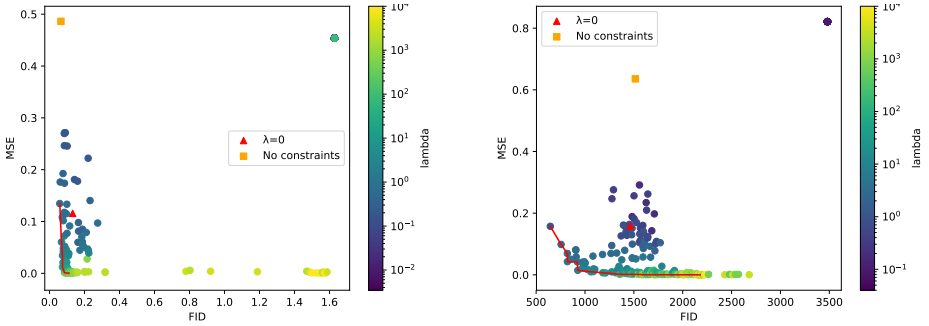


Fig. 5: MSE w.r.t the FID. Left: MNIST; Right: Fashion MNIST.

We evaluate our models based on both the satisfaction of the constraints and the visual quality of the generated samples. On one hand, we use the mean squared error between the provided constrained values and the constrained pixels in the generated image. On the other hand, evaluating the visual quality of an image is not a trivial task [13]. However, the recently developed metric referred to as Fréchet Inception Distance (FID) [14] seems to be a good metric of performance. Since using the FID requires a pre-trained classifier, we trained a simple convnet with MNIST/FashionMNIST labels as target. Lower layers of the classifier are then used to produce high-level features needed by the distance:

$$FID = ||\mu_r - \mu_g||2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}), \tag{5}$$

where μ_r, Σ_r, μ_g and Σ_g are the mean and the covariance matrices of extracted features obtained on respectively the real and the generated data. To overcome classical GANs instability, the networks are trained 10 times and the median of the best scores on the test set at the best epoch are recorded. The epoch that minimizes $\sqrt{FID^2 + MSE}$ on the validation set is considered as the best epoch.

Empirical evidences show that with a good choice of λ, the regularization term helps the generator to learn enforcing the constraints (Fig.3), leading to smaller MSEs than when using the CGAN approach only ($\lambda = 0$) and with minor detrimental effects on the quality of the samples (Fig.3). For Fashion MNIST, the regularization term even leads to a better image quality compared to the quality provided by GAN and CGAN approaches. Fig. 5 illustrates that the trade-off between image quality and the satisfaction of the constraints can be controlled by appropriately setting the value of λ. Nevertheless, for small values of λ the GAN fails to learn and only generates completely black samples. This leads to the plateaus seen for both the MSE and the FID in Fig. 3.

5 Conclusion

In this paper, we investigate the effectiveness of adding a regularization term to the conditioning of GANs to deal with cases where only a small subset of

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

the image one wants to generate is known beforehand. Empirical evidences illustrate that the proposed framework helps obtaining good image quality while best fulfilling the constraints compared to classical GAN approaches. In future work, we plan to extend this study to GAN conditioning in situations where no trivial mapping exists between the conditions and the generated samples, such as class-wise conditioning or more structured conditions.

References

[1] Eric Laloy, Romain Hérault, Diederik Jacques, and Niklas Linde. Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, 54(1):381–406, 2018.

[2] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[3] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[4] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[6] Raymond A Yeh, Chen Chen, Teck-Yian Lim, Alexander G Schwing, Mark Hasegawa-Johnson, and Minh N Do. Semantic image inpainting with deep generative models. In *CVPR*, volume 2, page 4, 2017.

[7] Lukas Mosser, Olivier Dubrule, and Martin J Blunt. Conditioning of three-dimensional generative adversarial networks for pore and reservoir-scale models. *arXiv preprint arXiv:1802.05622*, 2018.

[8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2017.

[9] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. *arXiv preprint arXiv:1801.07892*, 2018.

[10] Ugur Demir and Gozde Unal. Patch-based image inpainting with generative adversarial networks. *arXiv preprint arXiv:1803.07422*, 2018.

[11] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[12] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[13] Lucas Theis, Aäron van den Oord, and Matthias Bethge. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*, 2015.

[14] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017.