Normandie Université

# THÈSE

**Pour obtenir le grade de Docteur de Normandie Université**

**Spécialité Informatique**

**l'École Doctorale Mathématiques, Information, Ingéniérie des Systèmes**

# Auxiliary Tasks for the Conditioning of Generative Adversarial Networks

**Tâches auxilliaires pour le conditionnement des réseaux antagonistes génératifs**


**Présentée et soutenue par**

**Cyprien RUFFINO**

**Dirigée par Gilles GASSO et Romain HÉRAULT**


**Thèse soutenue publiquement le Monday 4th May, 2020**
**devant le jury composé de**

| | | |
|---|---|---|
| Civilité / prénom NOM, | Grade / fonction / statut / lieu d'exercice | Rapporteur ou examinateur ou directeur de thèse ou codirecteur de thèse |
| Civilité / prénom NOM, | Grade / fonction / statut / lieu d'exercice | Rapporteur ou examinateur ou directeur de thèse ou codirecteur de thèse |
| Civilité / prénom NOM, | Grade / fonction / statut / lieu d'exercice | Rapporteur ou examinateur ou directeur de thèse ou codirecteur de thèse |
| Civilité / prénom NOM, | Grade / fonction / statut / lieu d'exercice | Rapporteur ou examinateur ou directeur de thèse ou codirecteur de thèse |
| Civilité / prénom NOM, | Grade / fonction / statut / lieu d'exercice | Rapporteur ou examinateur ou directeur de thèse ou codirecteur de thèse |

INSA ROUEN NORMANDIE

MIΣ

litis

# Abstract

# Résumé

# Remerciements

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| CGAN | Conditional Generative Adversarial Networks |
| CycleGAN | Cycle-Consistent Generative Adversarial Networks |
| ELBO | Evidence Lower Bound |
| FID | Fréchet Inception Distance |
| GAN | Generative Adversarial Networks |
| GMM | Gaussian Mixture Model |
| IS | Inception Score |
| JS | Jensen-Shannon (Divergence) |
| KL | Kullback-Leibler (Divergence) |
| MSE | Mean-Squared Error |
| ReLU | Rectified Linear Unit |
| VAE | Variational Auto-Encoder |

# Introduction

## Context

Generic deep learning introduction, generic introduction to generative modeling (image generation, whichfaceisreal.com, etc...)

Generative Adversarial Networks ( GAN ) [1] have been recently highlighted for their ability to generate photo-realistic images. By providing a simple framework for high-quality, high-dimensional generative modeling, they quickly found real-world applications ranging from the notorious "deepfakes" [2] to image super-resolution [**?**].

Introduction to applied conditional generative modeling : examples others than geology

## Motivations

Geostatistical application : introduction and needs

- Tuneable (quality vs enforcement of the constraints)
- Pixel-precise
- Keeping diversity

Polarimetry application : introduction and needs

- Designing custom-made constraints for the problem
- Non-euclidian
- Compatible with domain transfer

## Contributions

## Outline

# Chapter 1

# Generative Adversarial Networks : Principles, strengths and limitations


### *Chapter abstract*

*In this chapter, we first propose an introduction to the problem of generative modeling and some solutions to tackle this problem. We then propose an overview of the Generative Adversarial Networks [1] framework, which is a recent method to train deep neural networks as generative models that is particularly adapted to the task of image generation. We will introduce some of its theoretical interpretations, as well as some of its variations and applications. We discuss the different limitations of this approach and expose a trilemma between the quality of the generated samples, their diversity and the conditioning of the model. We then discuss the recent advances that have been made to overcome some of these limitations and propose a taxonomy of these advances using the aforementioned trilemma. Finally, we discuss the evaluation of generative models and the difficulties of evaluating the intrinsic quality of a generated sample. We propose an overview of the different classical metrics and discuss their limitations.*


**Contents**

## 1.1 Generative modeling

we will first propose an overview of generative modeling

Generative modeling with deep neural networks has been a challenging task due to the stochastic nature of sampling, which prevents the computation of gradient, thus preventing the training of a deep model with gradient descent. CR: TODO

### 1.1.1 Generative modeling with maximum likelihood estimation

Generative modeling is the task of learning the underlying distribution of a dataset in order to generate more samples from that distribution. In other words, it describes how data is generated in terms of a probabilistic model, a distribution from which the whole dataset could have been sampled with a high likelihood.

Indeed, whereas a discriminative model tries to find decision boundaries by fitting a parametric model $p_\theta(y|x)$ to a conditional probability distribution $p(y|x)$ of labels $y \in \mathcal{Y}$ relatively to the data $x \sim p(x)$, a generative model aims to fit $p_\theta(x)$ to $p(x)$ the intrinsic distribution of the data and to provide a sampling mechanism on this distribution (see Figure. 1.1).

These two learning tasks, the discriminative (Equation. (1.1)) modeling and the generative modeling ( Equation. (1.2)) can be formulated as a maximum log-likelihood estimation[?]

$$\theta^* = \arg\max_\theta \mathop{\mathbb{E}}_{x,y\sim p(y|x)} \log p_\theta(y|x) \tag{1.1}$$

$$\theta^* = \arg\max_\theta \mathop{\mathbb{E}}_{x\sim p(x)} \log p_\theta(x) \tag{1.2}$$

An simple example of generative model are Gaussian Mixture Models (GMM) . They consist in a sum of K Gaussian distributions $\mathcal{N}(\mu_k, \sigma_k^2)$, $k \in 1..K$ which are all attributed a selection probability $p(z = k) = \pi_k$, so that $p(x|z = k) = \mathcal{N}(\mu_k, \sigma_k^2)$ . The model is then formulated as

$$p_\theta(x) = \sum_z p(z)p_\theta(x|z) \quad ,$$

with log-likelihood

$$\log \sum_{x\sim p(x)} p_\theta(x) = \sum_{x\sim p(x)} \log \sum_{k=1}^{k} \pi_k \mathcal{N}(x|\mu_k, \sigma_k^2) \quad .$$

In the case of the GMMs, the Expectation-Maximization (EM) algorithm [3] can be used to find the parameters $\theta^*$ which, at convergence, maximizes the log-likelihood of the model. Once the model is trained, sampling a new data is done by picking a component $k$ from the distribution $p(z)$, then drawing a sample from the Gaussian distribution $p(x|z = k) = \mathcal{N}(\mu_k^*, \sigma_k^{*2})$.

### 1.1.2 Latent variable models

#### Latent variable models and marginalization

For GMMs, sampling a new point consists in, once the Gaussian component has been selected, sampling a point on a normal distribution. This sampling can be done by using reparametrization: instead directly sampling $x \sim \mathcal{N}(\mu_k^*, \sigma_k^{*2})$, we can instead sample a latent variable $z \sim \mathcal{N}(0, 1)$ and compute $x = G(z; \mu, \sigma) = \mu + z\sigma$. Such a model, that consists in a deterministic function $G : \mathcal{Z} \to \mathcal{X}$ with parameters $\theta$ applied to a random latent variable drawn from a fixed distribution $p(z)$ is a latent variable model.

Since more complex distributions does not necessarily provide a natural sampling mechanism, using a latent variable model allows to outsource the stochastic part of the sampling process from the learning process and only learn the function $G(z; \theta)$. More formally, instead of directly modeling $p(x)$, a latent variable model learns a deterministic mapping $p_G(x|z)$. From this mapping, the generative model can be obtain through marginalization

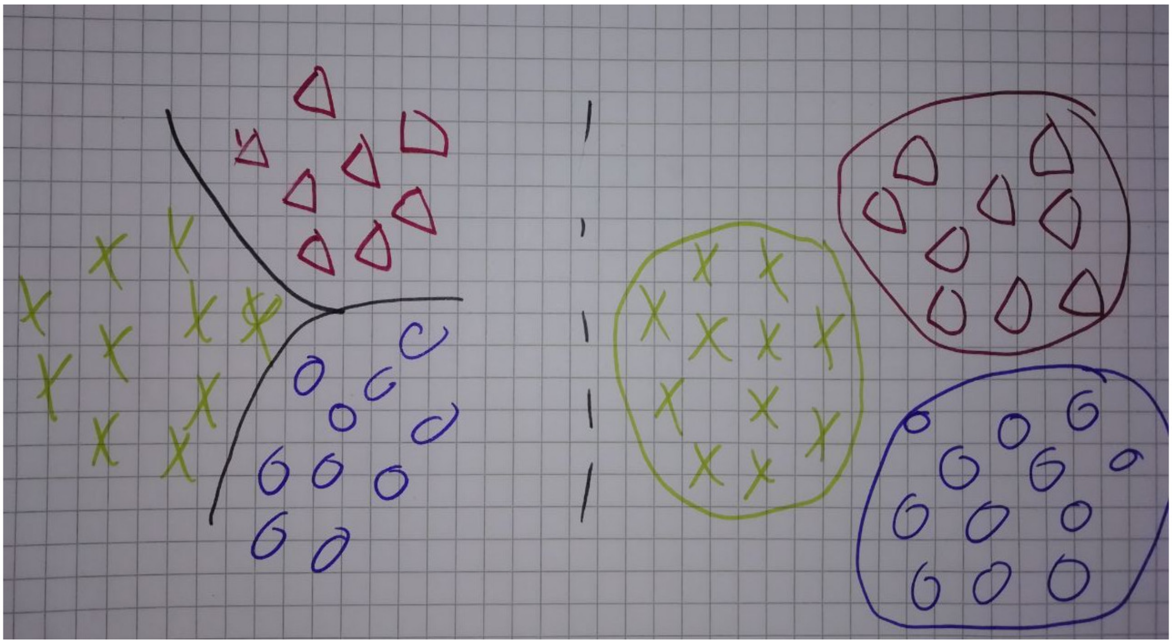


Figure 1.1: Left: Discriminative modeling, the model aims to find decision boundaries between classes. Right: Generative modeling, the model aims to learn the probability distribution of the data.

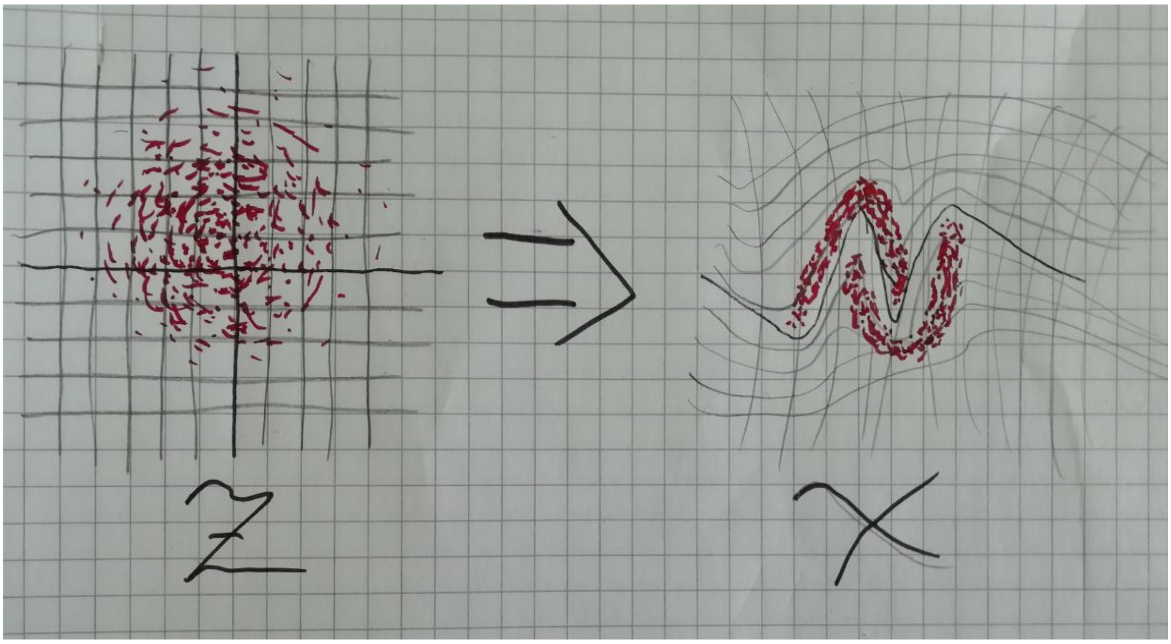



Figure 1.2: A mapping between a latent space $\mathscr{Z}$ and the space of a dataset $\mathscr{X}$.

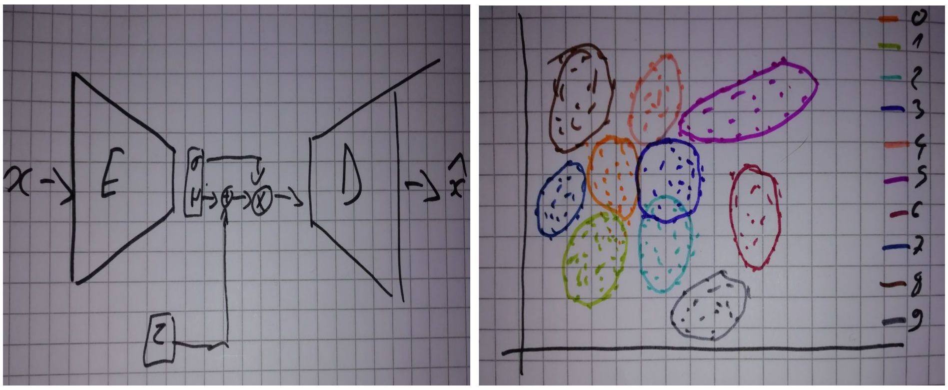



Figure 1.3: Left: the Variational auto-encoder global architecture. Right: the latent space of a VAE trained on the MNIST [5] dataset. We can observe a Gaussian component per class in the latent space,

$$p_{\mathrm{G}}(x) = \int_{\mathcal{Z}} p(z) p_{\mathrm{G}}(x|z) dz = \int_{\mathcal{Z}} p(z) p(x|\mathrm{G}(z;\theta)) dz \ . \tag{1.3}$$

This marginalization allows for the use of an arbitrary flexible G. However, if G is non-linear, the actual evaluation of $p_{\mathrm{G}}(x)$ is very likely to be intractable due to the integral over $\mathcal{Z}$, which prevents the training of such a model as is.

While the marginal distribution $p_{\mathrm{G}}(x)$ cannot be explicitly computed for any function G, several solutions exist to overcome this problem and train deep generative models with latent variables anyways.

### Variational auto-encoders

Variational Auto-Encoders (VAE) [4] are deep latent variable models which tackle the marginalization problem by approximating the integral using a variational approach. To this end, they both learn the distribution of the latent model $p_{\mathrm{G}}(x|z)$ as well as the distribution $q_{\mathrm{F}}(z|x)$. This is done with with two different neural networks, a decoder network $\mathrm{G}: \mathcal{Z} \to \mathcal{X}$ and an encoder network $\mathrm{F}: \mathcal{X} \to \mathcal{Z}$ and allows to compute the distribution $p(x)$ as

$$\log p_{\mathrm{G}}(x) - \mathrm{D_{KL}}\Big(q_{\mathrm{F}}(z|x)\Big|\Big|p(z|x)\Big) = \mathop{\mathbb{E}}_{z\sim q_{\mathrm{F}}(z|x)} \left[\log p_{\mathrm{G}}(x|z)\right] - \mathrm{D_{KL}}\Big(q_{\mathrm{F}}(zx)\Big|\Big|p(z)\Big) \ .$$

The KL terms evaluates the distance between the distribution $q_{\mathrm{F}}(z|x)$ learned by the encoder and real distribution $p(z|x)$, and since $p(z)$ is chosen Gaussian, this KL terms can be explicitly computed. The first term, is equivalent to the reconstruction error of an auto-encoder. Hence the model is trained by minimizing

$$\mathrm{L_{VAE}}(\mathrm{F},\mathrm{G}) = \mathop{\mathbb{E}}_{z\sim q_{\mathrm{F}}(z|x)} \left[||x - \mathrm{G}(z)||_2^2\right] - \mathrm{D_{KL}}\Big(q_{\mathrm{F}}(z|x)\Big|\Big|p(z)\Big)$$

However, since sampling $z \sim q_{\mathrm{F}}(z|x)$ is not differentiable, the VAE uses the so-called *reparametrization trick*, that is to have $\mathrm{F}(x)$ output the mean and the variance $(\mu_x, \sigma_x^2)$ of a normal distribution for a sample $x$, so that a $z' \sim \mathcal{N}(0,1)$ is sampled outside of the model and given as a parameter, thus allowing to compute $z = \mu_x + \sigma_x z'$, which is differentiable by considering $z'$ a parameter.

Finally, generating a sample $x$ with the trained model can be done by sampling a random vector $z \sim \mathcal{N}(0,1)$ and computing $x = \mathrm{G}(z)$.

### Normalizing flows

Normalizing flow based techniques is a family of latent variable models that aim to tackle the marginalization problem by using the *change of variable formula*

$$p_G(x) = p(z)\left|\det\left(\frac{\partial G(z)}{\partial z^T}\right)\right|^{-1} = p(G^{-1}(x))\left|\det\left(\frac{\partial G^{-1}(x)}{\partial x^T}\right)\right| ,$$

with $z \sim p(z)$ a latent variable. This formulation has notable advantages such as explicitly allowing the computation of the exact inference. However, the model has to enforce some tough constraints: the input and output dimensions must be the same; G must be invertible; and computing the determinant of the Jacobian needs to be efficient and differentiable.

These constraints can be enforced through strong restrictions on the architecture of the model. By limiting the transformations to a set of invertible transformations with a tractable Jacobian determinant, the model remains invertible and the determinant of its Jacobian can be computed efficiently.

Real-valued non-volume preserving (RealNVP) normalizing flows [**?**] uses affine coupling transformations, which transforms a variable by adding and scaling it by a non-linear transformation of itself, usually computed with deep neural networks. These transformations can be inverted by substracting and downscaling by the same transformed variables and their Jacobian is triangular, therefore computing its determinant can be done efficiently by computing the product of its diagonal terms. *Glow* [6] extended this set of transformations to $1 \times 1$ invertible convolutions as well as a variant of batch normalization that allows for more expressiveness in the model.

## 1.2 Generative Adversarial Networks

In this section, we will focus on the Generative Adversarial Networks [1] framework, their training process and some of their variants, especially their conditional and domain-transfer variants.

We will then outline some limitations of this framework and propose a formulation of these limitations in the form of a trilemma between the intrinsic quality of the generated samples, their diversity and the quality of the conditioning of the model. With this tool, we propose a taxonomy of the recent GAN approaches and identify trends in these approaches.

### 1.2.1 The GAN framework

In the same fashion as the generative models mentioned in Subsection. 1.1.2, Generative Adversarial Networks (GANs)[1] aims to learn a parameterized mapping $p_G(x|z)$ between a simple distribution $p(z)$ (usually normal or uniform) to the real data distribution $p(x)$. However, instead of relying on the likelihood and trying to estimate the distribution through marginalization, it aims to minimize an estimation of a divergence between $p(x)$ and the mapped distribution $p_G(x)$.

A divergence $\mathrm{Div}(p(x)||q(x))$ between two distributions $p(x)$ and $q(x)$ is analogous to a distance between these distributions. Thus, minimizing such a divergence allows for a parametric distribution $p_\theta(x)$ to converge to a target distribution $p(x)$. When this divergence is both tractable (or estimable) and differentiable w.r.t the parameters $\theta$, it can be directly optimized, allowing for the training of a generative model.

In practice, divergences are usually intractable. GANs estimate this divergence by relying on a second learned function that will act as an adversary: the discriminator D. The discriminator is a binary classifier that aims to predict the probability that a sample $x$ was sampled on the real distribution $p(x)$ or was generated from $z \sim p(z)$ and is trained by minimizing the binary cross-entropy. The objective of the generator G is then to fool the discriminator by maximizing the same binary cross-entropy. This training process is summed up as a mini-max game in Equation. (1.4)

$$\arg\min_G \max_D L_{GAN} = \arg\min_G \max_D \mathop{\mathbb{E}}_{x\sim p(x)}[\log D(x)] + \mathop{\mathbb{E}}_{z\sim p(z)}[1 - \log D(G(z))] \ . \tag{1.4}$$

This mini-max game has, assuming infinite capacity for both G and D, a global optimum for $p(x) = p_G(x)$, which can be proved as follows: the minimum of $f(x) = a\log(x) + b\log(1-x)$ is $\frac{a}{a+b}$, the discriminator that maximizes the criterion for a fixed G is given by

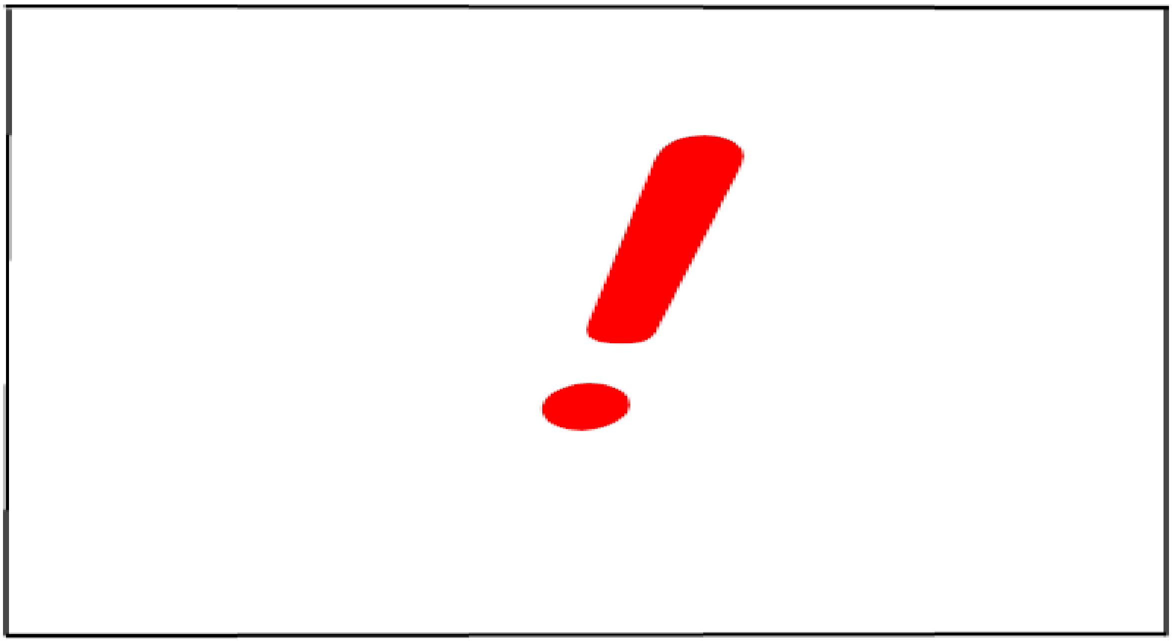


Figure 1.4: Generative Adversarial Networks framework

$$D^*_G(x) = \frac{p(x)}{p(x) + p_G(x)} \quad .$$

We can formulate a criterion C(G) by plugin this optimal discriminator in Equation. (1.4)

$$\begin{aligned}
C(G) &= \max_D L_{GAN}(D, G) \\
&= \mathop{\mathbb{E}}_{x\sim p(x)} \left[\log D^*(x)\right] + \mathop{\mathbb{E}}_{z\sim P_z} \left[1 - \log D^*(G(z))\right] = \mathop{\mathbb{E}}_{x\sim p(x)} \left[\log D^*(x)\right] + \mathop{\mathbb{E}}_{x\sim p_G(x)} \left[1 - \log D^*(x)\right] \\
&= \mathop{\mathbb{E}}_{x\sim p(x)} \left[\log \frac{p(x)}{p(x) + p_G(x)}\right] + \mathop{\mathbb{E}}_{x\sim p_G(x)} \left[1 - \log \frac{p_G(x)}{p(x) + p_G(x)}\right] \quad .
\end{aligned}$$

Up to additive and multiplicative constants, the criterion C(G) can be reformulated as

$$C(G) = D_{KL}\left(p(x) \Big|\Big| \frac{p(x) + p_G(x)}{2}\right) + D_{KL}\left(p_G(x) \Big|\Big| \frac{p(x) + p_G(x)}{2}\right) = 2 \cdot D_{JS}\left(p(x) \Big|\Big| p_G(x)\right) \quad .$$

When the discriminator is trained to convergence, minimizing the criterion $C(G) = L_{GAN}(D^*, G)$ is equivalent to minimizing the Jensen-Shannon (JS) divergence between $p(x)$ and $p_G(x)$. The GAN training process then consists in alternatively updating the discriminator and the generator via gradient ascent/descent. A summary of this process is presented in Algorithm. 1.

### 1.2.2 Conditional modeling with CGANs

While classical generative models such as GANs try to unconditionally approximate the real-data distribution $p(x)$, a conditional generative model aim to learn a model of the conditional distribution $p(x|y)$, where $y \in \mathcal{Y}$ is a label of any kind.

Several extensions of the GAN framework allow for conditional modeling. First introduced, Conditional GANs (CGANs)[1][7] simply adds the label $y$ as an input for both the discriminator and the generator. The new optimization problem that results from this change is summed-up in Equation. (1.5) as follows

---

**Algorithm 1** The GAN training algorithm

---

**Require:** $\mathscr{D}_X$ the real dataset, G the generator model, and D the discriminator model
**repeat**
  sample a mini-batch $\{x_i\}_{i=1}^m$ from $\mathscr{D}_X$
  sample a mini-batch $\{z_i\}_{i=1}^m$ from $p(z)$
  update D by stochastic gradient ascent of
    $\sum_{i=1}^m \log(D(x_i)) + \log(1 - D(G(z_i)))$
  sample a a mini-batch $\{z_j\}_{j=1}^n$ from distribution $p(z)$ ;
  update G by stochastic gradient descent of
    $\sum_{j=1}^n \log(1 - D(G(z_j)))$
**until** a stopping condition is met

---

$$\arg\min_G \max_D L_{CGAN} = \arg\min_G \max_D \mathop{\mathbb{E}}_{x,y\sim p(x,y)}[\log D(x,y)] + \mathop{\mathbb{E}}_{\substack{y\sim p(y)\\ z\sim p(z)}}[1 - \log D(G(y,z),y)] \tag{1.5}$$

While this approach is trivially simple to implement, it relies entirely on the discriminator to use the label. Other approaches try to learn the conditional distribution by adding an explicit loss term to the optimization problem, such as Auxillary Classifier GAN (ACGAN) [8]. This approach aims to learn a conditional generative model with discrete labels by adding another output to the discriminator that acts as a classifier. The model is then trained by having both the generator and the discriminator minimize the categorical cross-entropy between the real and predicted labels.

### 1.2.3 Domain-transfer with GANs

Domain-transfer is the task of learning a mapping $p(x|y)$ between two high-dimensional distributions $p(x)$ and $p(y)$ that maintains semantic information, for example changing the color palette of an image while keeping the same objects at the same position. CGANs already learn to model the conditional distribution $p(x|y)$, and adding a way to enforce the consistency of the semantic information enables domain-transfer.

Pix2Pix [9] implemented this approach explicitly by using paired samples $(x,y) \sim p(x|y)$ forcing the generator to minimize the $\ell_1$ reconstruction term between $x$ and $G(y,z)$ (Equation. (**??**)).

$$\arg\min_G \max_D L_{p2p} = \arg\min_G \max_D L_{CGAN}(D,G) + \lambda \mathop{\mathbb{E}}_{\substack{x\sim p(x)\\ y\sim p(y)\\ z\sim p(z)}} ||x - G(y,z)||_1 \tag{1.6}$$

However, this kind of approaches rely on paired data which can be very hard to obtain, especially in the case of natural images. When trying for example to transfer images of zebras to images of horses, you need a dataset of very similar zebras and horses in the exact same position for the $\ell_1$ term to work.

This problem of paired data was solved by CycleGAN [10] using cycle-consistency. Instead of training a single model G with reconstruction between $x$ and $G(y,z)$, the CycleGAN approach train two domain-transfer models simultaneously: $G_{YX}$ and $G_{XY}$ that map samples from $p(y)$ onto $p(x)$ and $p(x)$ onto $p(y)$, respectively (see Figure. 1.5). This allows to compute the $\ell_1$ reconstruction errors $||x - G_{YX}(G_{XY}(x))||_1$ and $||y - G_{XY}(G_{YX}(y))||_1$, thus completely removing the need for paired data $(x,y)$. The training of the two models in done an adversarial setup, with two discriminators $D_X$ and $D_Y$, and is summed up as an optimization problem in Equation. (1.7)

$$\begin{aligned}\arg\min_{G_{XY},G_{YX}} \max_{D_X,D_Y} L_{CycGAN} = \arg\min_{G_{XY},G_{YX}} \max_{D_X,D_Y} & L_{GAN}(D_X,G_{YX}) + L_{GAN}(D_Y,G_{XY}) \\ & + \lambda \mathop{\mathbb{E}}_{x\sim p(x)} ||x - G_{YX}(G_{XY}(x)||_1 + \lambda \mathop{\mathbb{E}}_{y\sim p(y)} ||y - G_{XY}(G_{YX}(y)||_1\end{aligned} \tag{1.7}$$

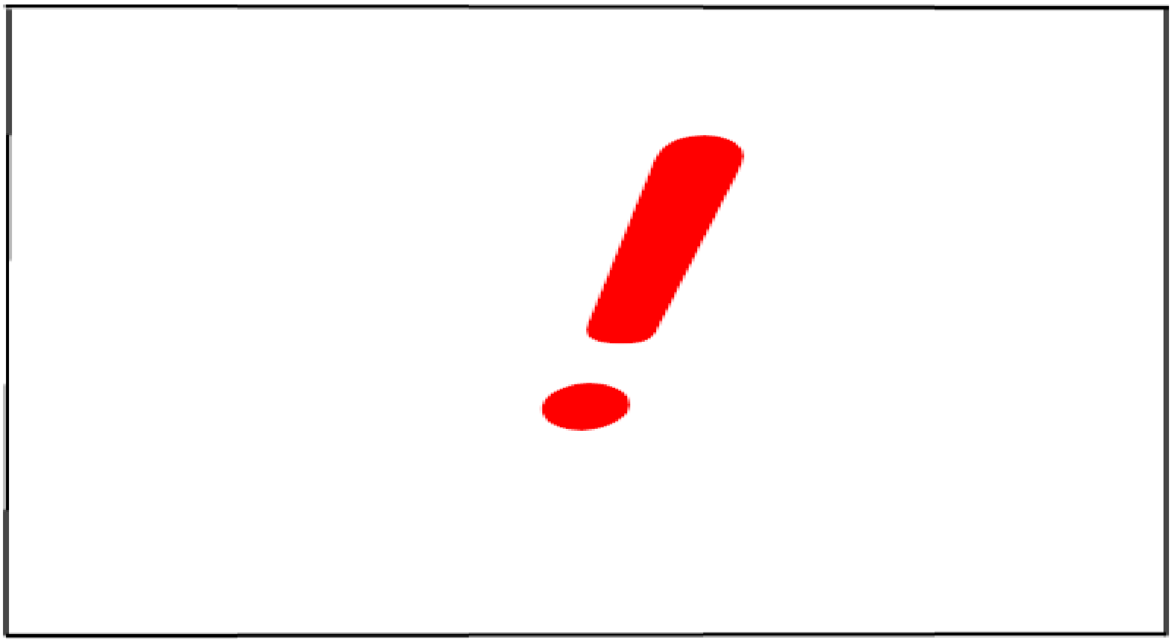


Figure 1.5: CycleGAN framework

The CycleGAN training process then consists in alternatively updating the two discriminator and the two generators via gradient ascent/descent. A summary of this process is presented in Algorithm. 2.

---

**Algorithm 2** CycleGAN training algorithm

---

**Require:** $\mathscr{X}$ and $\mathscr{Y}$ two unpaired datasets, $G_{XY}$ and $G_{YX}$ the mapping networks, $D_X$ and $D_Y$ the discrimination models, $m$ the mini-batch size

**repeat**

sample a mini-batch $\{x_i\}_{i=1}^{m}$ from $\mathscr{X}$

sample a mini-batch $\{y_i\}_{i=1}^{m}$ from $\mathscr{Y}$

update $D_X$ by stochastic gradient descent of

$$\sum_{i=1}^{m} (D_X(x_i) - 1)^2 + (D_X(G_{YX}(y_i)))^2$$

update $D_Y$ by stochastic gradient descent of

$$\sum_{i=1}^{m} (D_Y(y_i) - 1)^2 + (D_Y(G_{XY}(x_i)))^2$$

sample a mini-batch $\{x_i\}_{i=1}^{m}$ from X

sample a mini-batch $\{y_i\}_{i=1}^{m}$ from Y

update $G_{XY}$ by stochastic gradient descent of

$$\sum_{i=1}^{n} (D_Y(G_{XY}(x_i)) - 1)^2 + \lambda(||x_i - G_{YX}(G_{XY}(x_i))||_1 + ||y_i - G_{XY}(G_{YX}(y_i))||_1)$$

update $G_{YX}$ by stochastic gradient descent of

$$\sum_{i=1}^{n} (D_X(G_{YX}(y_i)) - 1)^2 + \lambda(||x_i - G_{YX}(G_{XY}(x_i))||_1 + ||y_i - G_{XY}(G_{YX}(y_i))||_1)$$

**until** a stopping condition is met

---

## 1.3 Limitations

GANs have shown strong advantages over the classical generative modeling methods, such as generating sharper samples than VAEs or taking significantly less time to train and to generate a sample than normalizing flows. However, they exhibit limitations: the instability of the training process; the lack diversity of the generated samples (*mode-collapse*); and finally the problems due to conditioning.

The instability of the GAN training process has first been conjectured to be caused by the bad quality of the gradients obtained when G generates bad samples, which makes D strongly reject these samples and therefore saturating the loss. The first solution proposed [1] was to slightly change the generator's loss function from $\log(1-D(G(z)))$ to $-\log(D(G(z)))$, which helped considerably to avoid failures of the training process and was then widely used

While this loss term converges to the same minimum as the original loss term, it however no longer correspond to a JSD

CR: Image quality : Incremental enhancement through architecture, more data, ...

Instability, catastrophic forgetting and the mode collapse problem

Trade-off image quality/diversity : Explanation through the loss term and distribution coverage

Black-box approach to conditioning, no tuning possible, no interpretability

## 1.4 The GAN Zoo

### 1.4.1 A taxonomy of GANs

Enorme nombre de variantes de GANs

Taxonomie des approches GANs (pour s'éviter une liste des différents GANs)

Schéma pour définir les grandes familles de GAN (évoquer les AmbientGAN / UNIR)

### 1.4.2 Architecture variants

### 1.4.3 Divergence variants

Table des loss alternatives (f-divergences + transport optimal)

### 1.4.4 Task-specific losses

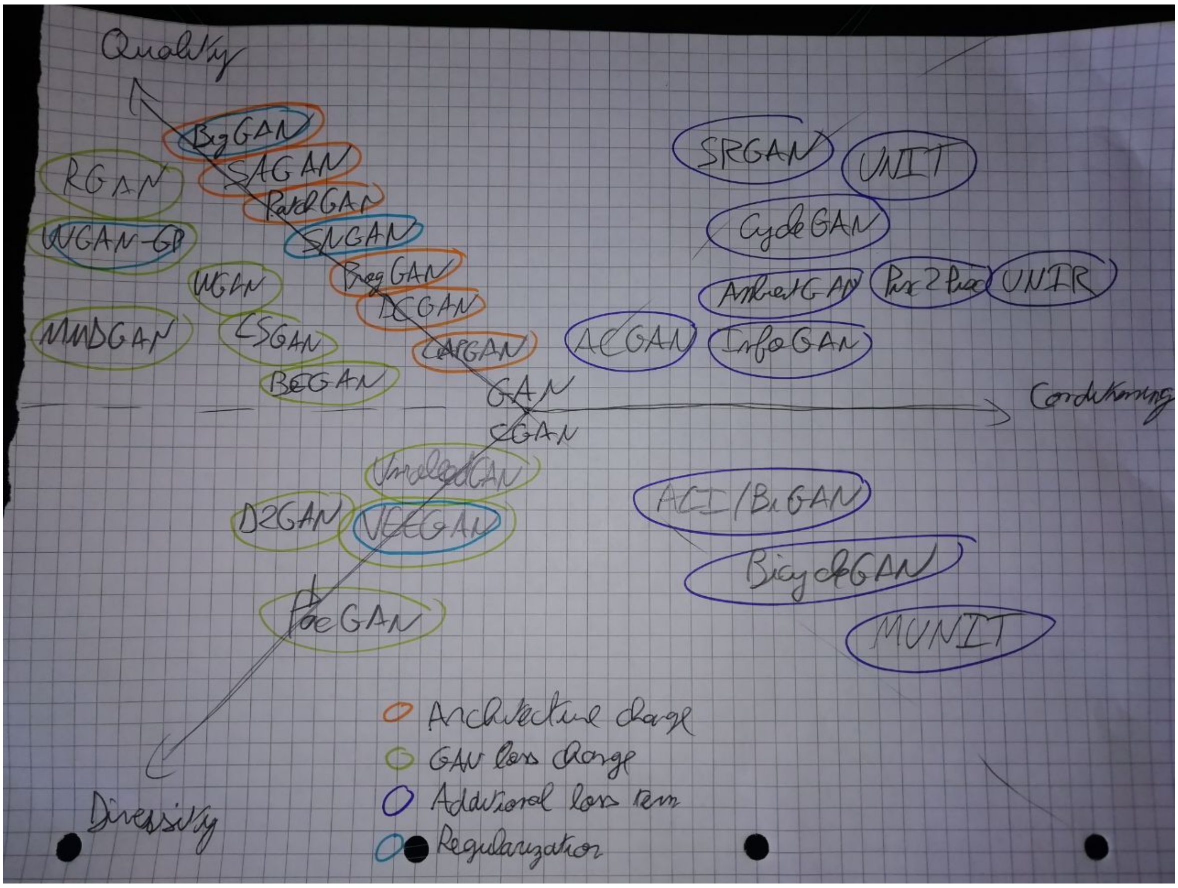



Figure 1.6: Classifications of some advances in GANs on the trilemma

## 1.5 A note on the evaluation of generative models

No good adhoc methods

Image quality : Inception distance + Fréchet inception distance, advantages

Conditioning : Direct evaluation (pixel-wise), Classifier accuracy, Projections (PCA, t-SNE)

Limitations of those metrics : need a pre-trained model

# Chapter 2

# Reconstruction as an Auxiliary Task for Generative Modeling


### *Chapter abstract*

*In this chapter, we propose an approach for conditioning a GAN model to reconstuct images from a very sparse set of randomly-positioned pixels known beforehand. This approach, based on a Maximum A Posteriori estimation, takes the form of an explicit auxiliary reconstruction task which adds to the GAN objective as an additional loss term. Complemented with the PacGAN variant for training GANs, this approach enables the generation of diverse samples from a sparse pixel map. As opposed to the more classical Conditional GAN approach, this auxiliary task is interpretable and a hyperparameter allows to control the importance of the conditioning in the learning process. We evaluate our approach on the classical MNIST, FashionMNIST and CIFAR10 datasets, as well as a custom-made texture dataset. Finally, we apply this approach to a task of geostatistical simulation.*


**Contents**



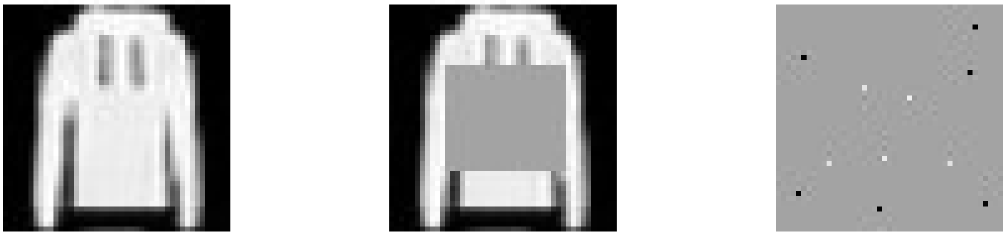
## 2.1 Image Reconstruction with Generative Models

Image reconstruction is the task of completing an image from a very small subset of the pixels. Such source data can usually be found in domains where the measurement process is very noisy or where measurements are expensive. This task differs from image inpainting since the source data is usually unstructured and very scarce, as in this chapter we will consider randomly scattered measurements of less than a percent of the image. While our discussion focus on image reconstruction, it is noteworthy to mention that this applies to other kinds of signals.

The task of image reconstruction is challenging since very few information is available for use. To overcome this lack of information, generative models such as GANs leverage on existing datasets to learn the distribution of the real images. By conditioning the learned distribution, a GAN could learn to generate an image while enforcing the constraint that the pixels known beforehand must remain similar in the generated image.

Similarly as in the GAN setup, we denote by $\mathrm{X} \in \mathscr{X}$ a random variable and $x$ its realization. Let $p_{\mathrm{X}}$ be the distribution of X over $\mathscr{X}$ and $p_{\mathrm{X}}(x)$ be its evaluation at $x$. Similarly $p_{\mathrm{X}|\mathrm{Y}}$ represents the distribution of X conditioned on the random variable $\mathrm{Y} \in \mathscr{Y}$.

We denote by $x \in \mathscr{X} = [-1,1]^{n\times p\times c}$ (see Figure 2.1a) an image sampled from an unknown distribution $p_{\mathrm{X}}$ and a sparse matrix $y \in \mathscr{Y} = [-1,1]^{n\times p\times c}$ (Figure 2.1c) as the given constrained pixels.



(a) Original Image | (b) Inpainting Input | (c) Constraint Map

Figure 2.1: Difference between regular inpainting (2.1b) and the problem undertaken in this work (2.1c) on a real sample (2.1a).

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

Figure 2.2: Generation of a sample during training. We first sample an image from a training set (2.2a) and we sample the constraints (2.2b) from it. Then our GAN generates a sample (2.2c). The constraints with squared error smaller than $\epsilon = 0.1$ are deemed satisfied and shown by green pixels in (2.2d) while the red pixels are unsatisfied.

The problem then consists in finding a generative model G with inputs $z$ (a random vector sampled from a known distribution $p_Z$ over the space $\mathcal{Z}$) and constrained pixel values $y \in [-1,1]^{n\times p\times c}$ that maps the distribution $p_Z$ onto the conditional distribution $p_{X|Y}$ of the real images given the constraints $y$ (see Figure 2.2).

CR: Related works : CGAN, AmbientGAN, UNIR, Compressed Sensing with Meta-Learning
Limitations of these models

## 2.2 Conditional generation as a Maximum A Posteriori estimation

Approche de l'article NeuCom :

Formulation as a Maximum A Posteriori Estimation, assumptions (normal error)

Construction of the loss term using bayes rule and least squares

PacGAN for keeping the diversity

## 2.3 Experimental evaluation and application to underground soil generation

Datasets : MNIST/FashionMNIST/CelebA/Texture

Evaluation : MSE/FID; Epoch selection criterion

Architectures : Appendix ? DCGAN + SGAN (encoder-decoder)

Results : visible trade-off, good fidelity overall

Application to hydro-geology : subsurface dataset

Evaluation : MSE/HOG+LBP

## 2.4 Conclusion

Objective reached : tuneable loss, pixel-wise, keeping diversity

Applications in hydro-geology : papier Eric

Future works : other distributions (modelling error using Laplacian, beta or Poisson distributions)

# Chapter 3

# Conditioning generation with multiple task-specific constraints

### *Chapter abstract*

*content...*

**Contents**



## 3.1 Introduction

Formulation as a constrained optimization problem
Reformulation of CycleGAN as a constrained optimization problem
Relaxation of the constraints
Ici, expérimenter sur des datasets artificiels ?

## 3.2 Proximal method for non-Euclidean output space

Travail sur le proximal ?
Envelope theorem application

## 3.3 Application to RGB to Polarimetric domain transfer

Introduction to polarimetry-specific physical constraints (briefly, no need to write a physics essay)
Reformulation as constraints on the output space
Relaxations : $L_2$ term + rectified term
Dataset, Evaluation
Experiments and results

## 3.4 Conclusion

Relaxation of the constraints works even when a lot of constraints are applied
The application to the polarimetric dataset works
Future works : using adapted metrics for the non-euclidean outspace X

# Chapter 4

# Conclusion and Perspectives

# Bibliography


[1] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. 2014. 1, 3, 7, 8, 11

[2] Cristian Vaccari and Andrew Chadwick. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media and Society*, 6(1), 2020. 1

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, sep 1977. 4

[4] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. 6

[5] Yann Lecun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 6

[6] Durk P. Kingma and Prafulla Dhariwal. Glow: Generative Flow with Invertible 1x1 Convolutions, 2018. 7

[7] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. 8

[8] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional Image Synthesis with Auxiliary Classifier GANs. 9

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-Image Translation with Conditional Adversarial Networks. 2016. 9

[10] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A Efros, and Berkeley Ai Research. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks Monet Photos. 9

[11] Sepp Hochreiter. The Vanishing Gradient Problem during learning Recurrent Neural Nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(2):107–116, 1997.

[12] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, and Russ Webb. Learning from Simulated and Unsupervised Images through Adversarial Training.

[13] Felix A Gers, Jurgen Schmidhuber, and Fred Cummins. Learning to Forget: Continual Prediction with LSTM. 1999.

[14] Guillaume Renton, Clement Chatelain, Sebastien Adam, Christopher Kermorvant, and Thierry Paquet. Handwritten text line segmentation using Fully Convolutional Network. *ICDAR Workshop on Machine Learning*, 2017.

[15] Scott Reed, Zeynep Akata, Santosh Mohan, Samuel Tenka, Bernt Schiele, and Honglak Lee. Learning What and Where to Draw.

[16] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. nov 2015.

[17] Eric Laloy, Romain Hérault, John Lee, Diederik Jacques, and Niklas Linde. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Advances in Water Resources*, 110:387–405, dec 2017.

[18] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015.

[19] Yann N Dauphin, Junyoung Chung, and Yoshua Bengio. RMSProp and equilibrated adaptive learning rates for non-convex optimization.

[20] Yann LeCun and Yoshua Bengio. Convolutional Networks for Images, Speech and Time Series. 1995.

[21] Anupama Ray, Sai Rajeswar, and Santanu Chaudhury. A hypothesize-and-verify framework for Text Recognition using Deep Recurrent Neural Networks. *13th International Confrence on Document Analysis and Recognition - ICDAR'15*, pages 936–940, 2015.

[22] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks.

[23] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, Dimitris Metaxas, Xiaogang Wang, and Dimitris Metaxas. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks.

[24] Joan Andreu Sanchez, Veronica Romero, Alejandro H Toselli, and Enrique Vidal. ICFHR2016 Competition on Handwritten Text Recognition on the READ Dataset. *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, pages 630–635, 2016.

[25] Nelson Morgan and Hervé Bourlard. An Introduction to Hybrid HMM/Connectionist Continuous Speech Recognition. 1995.

[26] Cyprien Ruffino, Romain Hérault, Eric Laloy, and Gilles Gasso. Dilated Spatial Generative Adversarial Networks for Ergodic Image Generation. may 2019.

[27] Alex Graves and Jurgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM networks. *Proceedings of the International Joint Conference on Neural Networks*, 4(July):2047–2052, 2005.

[28] Ayan Sinha, Zhao Chen, Vijay Badrinarayanan, and Andrew Rabinovich. Gradient Adversarial Training of Neural Networks.

[29] Theodore Théodore Bluche. Deep Neural Networks for Large Vocabulary Handwritten Text Recognition. 2015.

[30] Ilya Tolstikhin, Sylvain Gelly Google Brain Zürich, Olivier Bousquet Google Brain Zürich, Carl-Johann Simon-Gabriel, and Bernhard Schölkopf. AdaGAN: Boosting Generative Models. Technical report.

[31] Jefferey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

[32] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning Internal Representations by Error Propagation. 1985.

[33] Duhyeon Bang and Hyunjung Shim. Improved Training of Generative Adversarial Networks using Representative Features. 2018.

[34] Gwendoline De Bie, Gabriel Peyré, Marco Cuturi, and Google Brain. Stochastic Deep Networks. Technical report, 2018.

[35] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic Image Synthesis with Spatially-Adaptive Normalization. Technical report.

[36] Rie Johnson and Tong Zhang. Composite Functional Gradient Learning of Generative Adversarial Models. 2018.

[37] Wojciech Zaremba, Ilya Sutskever, Oriol Vinyals, and Google Brain. Recurrent Neural Network Regularization. 2015.

[38] John Duchi, Elad Hazan, and Yoram Singer. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization.

[39] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang. Generalization and Equilibrium in Generative Adversarial Nets (GANs). Technical report.

[40] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. One-shot Learning with Memory-Augmented Neural Networks.

[41] John J Hopfield. Neural Networks and Physical Systems with Emergent Collective Computational Abilities Neural networks and physical systems with emergent collective computational abilities (associative memory/parallel processing/categorization/content-addressable memory/f. *Biophysics*, 79:2554–2558, 1982.

[42] Alex Graves and Jurgen Schmidhuber. Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks. *Advances in Neural Information Processing Systems 21, NIPS'21*, pages 545–552, 2008.

[43] Jiezhang Cao, Yong Guo, Qingyao Wu, Chunhua Shen, Junzhou Huang, and Mingkui Tan. Adversarial Learning with Local Coordinate Coding. 2018.

[44] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models.

[45] Tijmen Tieleman and Geoffrey E Hinton. RMSProp : Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 2012.

[46] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. Generative Image Inpainting with Contextual Attention. jan 2018.

[47] William Fedus, Mihaela Rosca, Balaji Lakshminarayanan, Andrew M Dai, Shakir Mohamed, and Ian Goodfellow. Many Paths to Equilibrium: GANs Do Not Need to Decrease a Divergence At Every Step. 2017.

[48] Calvin Seward, Thomas Unterthiner, Urs Bergmann, Nikolay Jetchev, and Sepp Hochreiter. First Order Generative Adversarial Networks.

[49] Amélie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin Murphy. XGAN: Unsupervised Image-to-Image Translation for Many-to-Many Mappings. Technical report.

[50] Neale Ratzlaff and Li Fuxin. HyperGAN: A Generative Model for Diverse, Performant Neural Networks, jan 2019.

[51] Tu Dinh Nguyen, Trung Le, Hung Vu, and Dinh Phung. Dual Discriminator Generative Adversarial Nets. Technical report.

[52] Cyprien Ruffino, Romain Hérault, Eric Laloy, and Gilles Gasso. Pixel-wise Conditioned Generative Adversarial Networks for Image Synthesis and Completion. *Neurocomputing*, apr 2020.

[53] Yann LeCun, Leon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. Efficient BackProp. *Springer*, 1998.

[54] Shane Barratt and Rishi Sharma. A Note on the Inception Score. Technical report.

[55] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution.

[56] Christian Szegedy, Sergey Ioffe, Christian Szegedy, and Sergey Ioffe. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. feb 2015.

[57] Alexey Kurakin, Google Brain, Ian J Goodfellow, and Samy Bengio. ADVERSARIAL MACHINE LEARNING AT SCALE. 2017.

[58] Yi Wang, Xin Tao, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia. Image Inpainting via Generative Multi-column Convolutional Neural Networks, 2018.

[59] James Bergstra and Yoshua Bengio. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13:281–305, 2012.

[60] Sebastian Ruder. An overview of gradient descent optimization algorithms *. 2016.

[61] Mihaela Rosca, Balaji Lakshminarayanan, and David Warde-Farley Shakir Mohamed DeepMind. Variational Approaches for Auto-Encoding Generative Adversarial Networks. Technical report.

[62] 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.

[63] Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, Olivier Bousquet, and Google Brain. Are GANs Created Equal? A Large-Scale Study.

[64] Ian Goodfellow and Ian Goodfellow Openai. NIPS 2016 Tutorial: Generative Adversarial Networks. 2016.

[65] Anh Nguyen, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, Jason Yosinski, Yoshua Bengio, Alexey Dosovitskiy, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, and Jason Yosinski. Plug & Play Generative Networks: Conditional Iterative Generation of Images in Latent Space.

[66] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. The Numerics of GANs.

[67] Thomas Lucas, Corentin Tallec, Jakob Verbeek, and Yann Ollivier. Mixed batches and symmetric discriminators for GAN training. 2018.

[68] Sercan Ö Arık, Markus Kliegl, Rewon Child, Joel Hestness, Andrew Gibiansky, Chris Fougner, Ryan Prenger, and Adam Coates. Convolutional Recurrent Neural Networks for Small-Footprint Keyword Spotting.

[69] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151, jan 1999.

[70] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. aug 2017.

[71] Urs Bergmann, Zalandode Nikolay Jetchev, and Zalandode Roland Vollgraf ROLANDVOLLGRAF. Learning Texture Manifolds with the Periodic Spatial GAN.

[72] Sebastien Thomas, Thierry Paquet, Laurent Heutte, and Clement Chatelain. Un systeme generique d'extraction d'information dans des documents manuscrits non-contraints.

[73] Alex Tong Lin, Wuchen Li, Stanley Osher, Guido Montúfar, Guido Mont´ufar, and Mont´ Mont´ufar. WASSERSTEIN PROXIMAL OF GANS. Technical report.

[74] Yang Zhou, Huajie Shi, Dani Lischinski, Minglun Gong, Johannes Kopf, and Hui Huang. Analysis and Controlled Synthesis of Inhomogeneous Textures. *Computer Graphics Forum*, 36(2):199–212, may 2017.

[75] Sudipto Mukherjee, Himanshu Asnani, Eugene Lin, and Sreeram Kannan. ClusterGAN : Latent Space Clustering in Generative Adversarial Networks. Technical report.

[76] Kevin J Liang, Chunyuan Li, Guoyin Wang, and Lawrence Carin. Generative Adversarial Network Training is a Continual Learning Problem. nov 2018.

[77] Alex Graves. Generating Sequences With Recurrent Neural Networks. 2013.

[78] Yoshua Bengio. Practical Recommendations for Gradient-Based Training of Deep Architectures. 2012.

[79] Tong Che, Yanran Li, Ruixiang Zhang, R Devon Hjelm, Wenjie Li, Yangqiu Song, and Yoshua Bengio. Maximum-Likelihood Augmented Discrete Generative Adversarial Networks. *2017.*

[80] Kyunghyun Cho, Bart Van Merriënboer, and Dzmitry Bahdanau. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. 2014.

[81] Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Olivier Mastropietro, Alex Lamb, Martin Arjovsky, and Aaron Courville. ADVERSARIALLY LEARNED INFERENCE.

[82] Andrew Brock, Jeff Donahue, Karen Simonyan, Jeff Donahue Deepmind, and Karen Simonyan Deepmind. Large Scale GAN Training for High Fidelity Natural Image Synthesis. sep 2018.

[83] Christopher M Bishop. *Neural Networks for Pattern Recognition.* 1995.

[84] Dario Amodei, Rishita Anubhai, Eric Battenberg, Case Carl, Jared Casper, Bryan Catanzaro, Jingdong Chen, Mike Chrzanowski, Adam Coates, Greg Diamos, Erich Elsen, Jesse Engel, Linxi Fan, Christopher Fougner, Tony Han, Awni Hannun, Billy Jun, Patrick LeGresley, Libby Lin, Sharan Narang, Andrew Ng, Sherjil Ozair, Ryan Prenger, Jonathan Raiman, Sanjeev Satheesh, David Seetapun, Shubho Sengupta, Yi Wang, Zhiqian Wang, Chong Wang, Bo Xiao, Dani Yogatama, Jun Zhan, and Zhenyao Zhu. Deep-speech 2: End-to-end speech recognition in English and Mandarin. *Jmlr W&Cp*, 48:28, 2015.

[85] Kihyuk Sohn, Xinchen Yan, and Honglak Lee. Learning Structured Output Representation using Deep Conditional Generative Models.

[86] Irwan Bello, Barret Zoph, Vijay Vasudevan, and Quoc V Le. Neural Optimizer Search with Reinforcement Learning. 2017.

[87] Paul J. Werbos. Backpropagation Through Time: What Id Does and How to Do It. *Proceedings of the IEEE*, 78(10), 1990.

[88] Mahyar Khayatkhoei, Ahmed Elgammal, and Maneesh Singh. Disconnected Manifold Learning for Generative Adversarial Networks. Technical report.

[89] Hai Dai Nguyen, Anh Duc Le, and Masaki Nakagawa. Deep neural networks for recognizing online handwritten mathematical symbols. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 121–125. IEEE, nov 2015.

[90] Jesse Engel, Matthew Hoffman, and Adam Roberts. Latent Constraints: Learning to Generate Conditionally from Unconditionnal Generative Models.

[91] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. nov 2018.

[92] Matti Pietikäinen, Abdenour Hadid, Guoying Zhao, and Timo Ahonen. *Computer Vision Using Local Binary Patterns*, volume 40 of *Computational Imaging and Vision*. Springer London, London, 2011.

[93] Vaishnavh Nagarajan and J. Zico Kolter. Gradient descent GAN optimization is locally stable. jun 2017.

[94] Matthew D Zeiler. AdaDelta: An Adaptative Learning Rate Method.

[95] Shirsendu Sukanta Halder, Kanjar De, and Partha Pratim Roy. Perceptual Conditional Generative Adversarial Networks for End-to-End Image Colourization. nov 2018.

[96] Martin Arjovsky and Léon Bottou. TOWARDS PRINCIPLED METHODS FOR TRAINING GENERATIVE ADVERSARIAL NETWORKS.

[97] Aäron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WAVENET: A GENERATIVE MODEL FOR RAW AUDIO.

[98] L. Lemmens, B. Rogiers, M. Craen, E. Laloy, D. Jacques, and et al. Huysmans, D. Effective structural descriptors for natural and engineered radioactive waste confinement barrier, 2017.

[99] Tameem Adel, Zoubin Ghahramani, and Adrian Weller. Discovering Interpretable Representations for Both Deep Generative and Discriminative Models. 2018.

[100] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, and Ilya Sutskever. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets.

[101] Stefan Webb, Adam Golinski, Rob Zinkov, Siddharth Narayanaswamy, Tom Rainforth, Yee Whye Teh, and Frank Wood. Faithful Inversion of Generative Models for Effective Amortized Inference, 2018.

[102] Guim Perarnau, Joost Van De Weijer, Bogdan Raducanu, and Jose M Álvarez. Invertible Conditional GANs for image editing.

[103] Yichun Shi, Debayan Deb, and Anil K. Jain. WarpGAN: Automatic Caricature Generation. nov 2018.

[104] Piotr Bojanowski, Armand Joulin, David Lopez Paz, and Arthur Szlam. Optimizing the Latent Space of Generative Networks. 2018.

[105] Ryuhei Hamaguchi, Aito Fujita, Keisuke Nemoto, Tomoyuki Imaizumi, and Shuhei Hikosaka. Effective Use of Dilated Convolutions for Segmenting Small Object Instances in Remote Sensing Imagery. sep 2017.

[106] Theodore Bluche. Joint Line Segmentation and Transcription for End-to-End Handwritten Paragraph Recognition. pages 1–12, 2016.

[107] Yang Liu, Chengjie Sun, Lei Lin, and Xiaolong Wang. Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention. 2016.

[108] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian Optimization of Machine Learning Algorithms. 2017.

[109] Jesse Engel, Kumar Krishna Agrawal, Shuo Chen, Ishaan Gulrajani, Chris Donahue, and Adam Roberts. GANSYNTH: ADVERSARIAL NEURAL AUDIO SYNTHESIS. Technical report.

[110] Alex Graves, Santiago Fernandez, Jürgen Jurgen Schmidhuber, Santiago Fernández, and Jürgen Jurgen Schmidhuber. Multi-Dimensional Recurrent Neural Networks. 2007.

[111] Yann LeCun, Yoshua Bengio, and Geoffrey E Hinton. Deep learning. *Nature*, 521(7553):436–444, may 2015.

[112] Yuri Nesterov. A Method of Solving a Convex Programming Problem with Convergence Rate O(1/kˆ2). *Soviet Math Dokl.*, 27(2):372–376, 1983.

[113] Zecheng Xie, Zenghui Sun, Lianwen Jin, Ziyong Feng, and Shuye Zhang. Fully Convolutional Recurrent Network for Handwritten Chinese Text Recognition. 2016.

[114] Tycho FA van der Ouderaa and Daniel E Worrall. Reversible GANs for Memory-efficient Image-to-Image Translation. Technical report.

[115] Yang Zhou, Hui Huang, Zhen Zhu, Xiang Bai, Dani Lischinski, and Daniel Cohen-Or. Non-Stationary Texture Synthesis by Adversarial Expansion Additional Key Words and Phrases: Example-based texture synthesis, non-stationary textures, generative adversarial networks ACM Reference Format. *ACM Trans. Graph*, 37(13):13, 2018.

[116] Tomas Wilkinson and Anders Brun. Semantic and Verbatim Word Spotting using Deep Neural Networks. *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, 2016.

[117] Zhiting Hu, Zichao Yang, Ruslan R. Salakhutdinov, LIANHUI Qin, Xiaodan Liang, Haoye Dong, and Eric P. Xing. Deep Generative Models with Learnable Knowledge Constraints, 2018.

[118] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. An Empirical Exploration of Recurrent Network Architectures.

[119] Mehdi S M Sajjadi, Giambattista Parascandolo, Arash Mehrjou, and Bernhard Schölkopf. Tempered Adversarial Networks. 2018.

[120] Sharath Adavanne, Pasi Pertilä, and Tuomas Virtanen. SOUND EVENT DETECTION USING SPATIAL FEATURES AND CONVOLUTIONAL RECURRENT NEURAL NETWORK.

[121] Leifert Gundram, Tobias Strauß, Tobias Grüning, Welf Wustlich, and Roger Labahn. Cells in Multidimensional Recurrent Neural Networks. 2016.

[122] Ferenc Huszár. HOW (NOT) TO TRAIN YOUR GENERATIVE MODEL: SCHEDULED SAMPLING, LIKELIHOOD, ADVERSARY? 2015.

[123] Lingpeng Kong, Chris Alberti, Daniel Andor, Ivan Bogatyy, and David Weiss. DRAGNN: A Transition-based Framework for Dynamically Connected Neural Networks.

[124] Ugur Demir and Gozde Unal. Patch-Based Image Inpainting with Generative Adversarial Networks. mar 2018.

[125] Michael Arbel, Dougal J Sutherland, Mikołaj Bí, and Arthur Gretton. On gradient regularizers for MMD GANs. Technical report.

[126] Eric Laloy, Niklas Linde, Cyprien Ruffino, Romain Hérault, Gilles Gasso, and Diederik Jacques. Gradient-based deterministic inversion of geophysical data with generative adversarial networks: Is it feasible?, dec 2019.

[127] Diederik P Kingma, Danilo J Rezende, Shakir Mohamed, and Max Welling. Semi-supervised Learning with Deep Generative Models. Technical report.

[128] Anders Krogh and John A Hertz. A Simple Weight Decay Can Improve Generalization. 4:950–957, 1992.

[129] Xudong Pan, Mi Zhang, and Daizong Ding. Theoretical Analysis of Image-to-Image Translation with Adversarial Learning. 2018.

[130] Valentin Khrulkov and Ivan Oseledets. Geometry Score: A Method For Comparing Generative Adversarial Networks. 2018.

[131] Chongxuan LI, Max Welling, Jun Zhu, and Bo Zhang. Graphical Generative Adversarial Networks, 2018.

[132] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep Sparse Rectifier Neural Networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 2011.

[133] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. IEEE.

[134] Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, Adrià Puigdomènech Badia, Karl Moritz Hermann, Yori Zwols, Georg Ostrovski, Adam Cain, Helen King, Christopher Summerfield, Phil Blunsom, Koray Kavukcuoglu, and Demis Hassabis. Hybrid computing using a neural network with dynamic external memory. *Nature*, 538(7626):471–476, oct 2016.

[135] F Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain. *Psychological Review*, pages 65–386, 1958.

[136] Diederik P Kingma, Prafulla Dhariwal, and San Francisco. Glow: Generative Flow with Invertible 1×1 Convolutions. Technical report.

[137] François Chollet and Others. *Keras*. 2015.

[138] Na Lei, Kehua Su, Li Cui, Shing-Tung Yau, and David Xianfeng Gu. A Geometric View of Optimal Transportation and Generative Model.

[139] 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.

[140] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation ppt. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[141] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. 2013.

[142] Kuzman Ganchev, João Graça, Jennifer Gillenwater, and Ben Taskar. Posterior Regularization for Structured Latent Variable Models. Technical report, 2010.

[143] Joao Neto, Luis Almeida, Mike Hochberg, Ciro Martins, Luis Nunes, Steve Renals, and Tony Robinson. Speaker-Adaptation For Hybrid HMM-ANN Continuous Speech Recognition System. 1994.

[144] Dan Zhang and Anna Khoreva. PA-GAN: Improving GAN Training by Progressive Augmentation. Technical report.

[145] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local Neural Networks. Technical report.

[146] Zhen Zuo, Bing Shuai, Gang Wang, Xiao Liu, Xingxing Wang, Bing Wang, and Yushi Chen. Convolutional Recurrent Neural Networks: Learning Spatial Dependencies for Image Representation.

[147] Tomas Jakab, Ankush Gupta, Hakan Bilen, and Andrea Vedaldi. Unsupervised Learning of Object Landmarks through Conditional Image Generation, 2018.

[148] Thomas Plötz and Gernot A. Fink. Markov models for offline handwriting recognition: A survey. *International Journal on Document Analysis and Recognition*, 12(4):269–298, dec 2009.

[149] Lucas Theis, Aäron Van Den Oord, and Matthias Bethge. A NOTE ON THE EVALUATION OF GENERATIVE MODELS.

[150] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. Technical report.

[151] Chenyang Tao, Liqun Chen, Ricardo Henao, Jianfeng Feng, and Lawrence Carin. $\chi$ 2 Generative Adversarial Network. 2018.

[152] Songyao Jiang, Hongfu Liu, · Yue Wu, Yun Fu, and Yue Wu. Spatially Constrained Generative Adversarial Networks for Conditional Image Generation. Technical report.

[153] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved Training of Wasserstein GANs. 2017.

[154] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. 2012.

[155] Diederik P Kingma and Jimmy Lei Ba. Adam : A Method for Stochastic Optimization.

[156] Augustus Odena, Jacob Buckman, Catherine Olsson, Tom B Brown, Christopher Olah, Colin Raffel, and Ian Goodfellow. Is Generator Conditioning Causally Related to GAN Performance? 2018.

[157] Chang Xiao, Peilin Zhong, and Changxi Zheng. BourGAN: Generative Networks with Metric Embeddings, 2018.

[158] Alex Graves. Supervised Sequence Labelling with Recurrent Neural Networks. 2012.

[159] Nikolay Jetchev, Urs Bergmann, Roland Vollgraf, and Zalando Research. Texture Synthesis with Spatial Generative Adversarial Networks. 2017.

[160] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning Algorithm for Boltzmann Machines. *Cognitive Science*, 9:147–169, 1985.

[161] Mahdi Hamdani, Patrick Doetsch, and Hermann Ney. Improvement of Context Dependent Modeling for Arabic Handwriting Recognition. *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, 2014-Decem:494–499, 2014.

[162] Paul Voigtlaender, Patrick Doetsch, and Hermann Ney. Handwriting Recognition with Large Multidimensional Long Short-Term Memory Recurrent Neural Networks. 2016.

[163] Liangjian Chen, Shih-Yao Lin, Yusheng Xie, Hui Tang, Yufan Xue, Xiaohui Xie, Yen-Yu Lin, and Wei Fan. Generating Realistic Training Images Based on Tonality-Alignment Generative Adversarial Networks for Hand Pose Estimation. nov 2018.

[164] Xinyue Zhu, Yifan Liu, Zengchang Qin, and Jiahong Li. Emotion Classification with Data Augmentation Using Generative Adversarial Networks. Technical report, 2017.

[165] Felix A Gers, Nicol N Schraudolph, and Jürgen Jurgen Schmidhuber. Learning Precise Timing with LSTM Recurrent Networks. *Journal of Machine Learning Research*, 3:115–143, 2002.

[166] Emmanuel Augustin, Matthieu Carré, Emmanuèle Grosicki, Jean-Marie Brodin, and Edouard Geoffrois. RIMES evaluation campaign for handwritten mail processing. 2006.

[167] Yoshua Bengio, Nicolas Boulanger-Lewandowski, and Razvan Pascanu. Advances in Optimizing Recurrent Networks. 2012.

[168] Chongxuan Li, Kun Xu, Jun Zhu, and Bo Zhang. Triple Generative Adversarial Nets.

[169] Kenneth O Stanley and Risto Miikkulainen. The MIT Press Journals Evolving Neural Networks through Augmenting Topologies.

[170] Jonathan Fiscus. A Post-Processing System To Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction (ROVER). (February), 1997.

[171] Nikolay Jetchev and Zalando Research. The Conditional Analogy GAN: Swapping Fashion Articles on People Images.

[172] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which Training Methods for GANs do actually Converge? 2018.

[173] Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, Noah A Smith, and Marianas Labs. Transition-Based Dependency Parsing with Stack Long Short-Term Memory.

[174] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved Techniques for Training GANs. jun 2016.

[175] Mehdi S. M. Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing Generative Models via Precision and Recall. may 2018.

[176] Hang Gao, Zheng Shou, Alireza Zareian, Hanwang Zhang, and Shih-Fu Chang. Low-shot Learning via Covariance-Preserving Adversarial Augmentation Networks. Technical report.

[177] Andrew Y Ng. Feature selection, L1 vs L2 regularization and rotational invariance. In *Proceedings of the 21 st International Conference on Machine Learning*, Banff, Canada, 2004.

[178] Georg Ostrovski, Will Dabney, and Rémi Munos. Autoregressive Quantile Networks for Generative Modeling. 2018.

[179] Léon Bottou and Olivier Bousquet. The Tradeoffs of Large Scale Learning.

[180] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. 2010.

[181] Giambattista Parascandolo, Niki Kilbertus, Mateo Rojas-Carulla, and Bernhard Schölkopf. Learning Independent Causal Mechanisms. 2018.

[182] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning Long Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994.

[183] Klaus Greff, Rupesh Srivastava Kumar, Jan Koutník, Hkou Ch, Bas Steunebrink, and Jurgen Schmidhuber. LSTM: A Search Space Odyssey. 2015.

[184] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative Adversarial Networks: An Overview. 2017.

[185] Vincent Andrearczyk Supervisor and Paul F Whelan. Deep learning for texture and dynamic texture analysis. 2017.

[186] Yanhai Gan, Huifang Chi, Ying Gao, Jun Liu, Guoqiang Zhong, and Junyu Dong Ocean. PERCEPTION DRIVEN TEXTURE GENERATION.

[187] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. 2017.

[188] Fisher Yu and Vladlen Koltun. MULTI-SCALE CONTEXT AGGREGATION BY DILATED CONVOLUTIONS.

[189] Alex Graves. RNNLIB: A recurrent neural network library for sequence learning problems. \url{http://sourceforge.net/projects/rnnl/}.

[190] Michał Kozielski, Patrick Doetsch, Mahdi Hamdani, and Hermann Ney. Multilingual Off-line Handwriting Recognition in Real-world Images.

[191] Akash Srivastava, Lazar Valkov, Chris Russell, Michael U. Gutmann, and Charles Sutton. VEEGAN: Reducing Mode Collapse in GANs using Implicit Variational Learning, 2017.

[192] Zinan Lin, Ashish Khetan, Giulia Fanti, and Sewoong Oh. PacGAN: The power of two samples in generative adversarial networks. Technical report, 2018.

[193] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition.

[194] Aaron Van Den Oord, Yazhe Li, Igor Babuschkin, Karen Simonyan, Oriol Vinyals, Koray Kavukcuoglu, George Van Den Driessche, Edward Lockhart, Luis C Cobo, Florian Stimberg, Norman Casagrande, Dominik Grewe, Seb Noury, Sander Dieleman, Erich Elsen, Nal Kalchbrenner, Heiga Zen, Alex Graves, Helen King, Tom Walters, Dan Belov, and Demis Hassabis. Parallel WaveNet: Fast High-Fidelity Speech Synthesis. 2018.

[195] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the CoordConv solution. Technical report.

[196] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A Fast Learning Algorithm for Deep Belief Nets. 2006.

[197] George Cybenkot. Approximation by Superpositions of a Sigmoidal Function*. *Math. Control Signals Systems*, 2:303–314, 1989.

[198] Antreas Antoniou, Amos Storkey, and Harrison Edwards. DATA AUGMENTATION GENERATIVE ADVERSARIAL NETWORKS.

[199] Lukas J Mosser, Olivier Dubrule, and Martin J Blunt. Conditioning of three-dimensional generative adversarial networks for pore and reservoir-scale models. Technical report.

[200] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[201] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-Supervised Generative Adversarial Networks. nov 2018.

[202] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[203] René Vidal, Joan Bruna, Raja Giryes, and Stefano Soatto. Mathematics of Deep Learning. Technical report.

[204] Lukas Balles and Philipp Hennig. Dissecting Adam: The Sign, Magnitude and Variance of Stochastic Gradients.

[205] David Berthelot, Google Brain, Colin Raffel, Aurko Roy Google Brain, and Ian Goodfellow Google Brain. Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer. Technical report.

[206] Takuhiro Kaneko, Yoshitaka Ushiku, and Tatsuya Harada. Class-Distinct and Class-Mutual Image Generation with GANs. nov 2018.

[207] Sepp Hochreiter and Jurgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1–32, 1997.

[208] Ronald J Williams and David Zipser. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, 1(2):270–280, 1989.

[209] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2015.

[210] Leon Bottou and Yann LeCun. Global Training of Document Processing Systems using Graph Transformer Networks.

[211] Kelvin Xu, Jimmy Lei Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Salakhutdinov Ruslan, Richard S Zemel, and Yoshua Bengio. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2016.

[212] Ali Borji. Pros and Cons of GAN Evaluation Measures. 2018.

[213] Catherine Olsson, Surya Bhupatiraju, Tom Brown, Augustus Odena, Ian Goodfellow, and Google Brain. Skill Rating for Generative Models. Technical report.

[214] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[215] Takuhiro Kaneko, Yoshitaka Ushiku, and Tatsuya Harada. Label-Noise Robust Generative Adversarial Networks. nov 2018.

[216] Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Kate Saenko Umass, Lowell Lowell, and Trevor Darrell. Long-term Recurrent Convolutional Networks for Visual Recognition and Description.

[217] Kyunghyun Cho, Aaron Courville, and Yoshua Bengio. Describing Multimedia Content Using Attention-Based Encoder-Decoder Networks. 17(11):1875–1886, 2015.

[218] Keunwoo Choi, Gyorgy Fazekas, Mark Sandler, and Kyunghyun Cho. Convolutional recurrent neural networks for music classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2392–2396. IEEE, mar 2017.

[219] Mikołaj Bińkowski, Dougal J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. jan 2018.

[220] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.

[221] Ilya Kamenshchikov and Matthias Krauledat. Effects of Dataset properties on the training of GANs. nov 2018.

[222] Theodore Bluche, Jérôme Louradour, and Ronaldo Messina. Scan, Attend and Read: End-to-End Handwritten Paragraph Recognition with MDLSTM Attention. pages 1–10, 2016.

[223] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training Recurrent Neural Networks.

[224] Warren S Mcculloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *BULLETIN OF MATHEMATICAL BIOPHYSICS*, 5, 1943.

[225] Alex Graves, Santiago Fernandez, Faustino Gomez, and Jurgen Schmidhuber. Connectionist Temporal Classification : Labelling Unsegmented Sequence Data with Recurrent Neural Networks. *Proceedings of the 23rd international conference on Machine Learning*, pages 369–376, 2006.

[226] Alex Graves. Hierarchical Subsampling Networks. pages 109–131, 2012.

[227] Emre Ç Akır, Giambattista Parascandolo, Toni Heittola, Heikki Huttunen, and Tuomas Virtanen. Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection.

[228] Yoon Kim, Yacine Jernite, David Sontag, and Alexander M Rush. Character-Aware Neural Language Models.

[229] Cyprien Ruffino, Romain Hérault, Eric Laloy, and Gilles Gasso. Pixel-wise Conditioning of Generative Adversarial Networks. *ESANN 2019 - Proceedings, 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 25–30, nov 2019.

[230] Matthias Holschneider. On the wavelet transformation of fractal objects. *Journal of Statistical Physics*, 50(5-6):963–993, mar 1988.

[231] Volodymyr Mnih, Nicolas Heess, and Alex Graves. Recurrent models of visual attention. *Nips*, pages 1–9, 2014.

[232] Ishan Deshpande, Ziyu Zhang, and Alexander Schwing. Generative Modeling using the Sliced Wasserstein Distance. mar 2018.

[233] Shibani Santurkar, Ludwig Schmidt, and Aleksander M̨ Adry. A Classification-Based Study of Covariate Shift in GAN Distributions. 2018.

[234] Zhijie Deng, Hao Zhang, Xiaodan Liang, Luona Yang, Shizhen Xu, Jun Zhu, and Eric P. Xing. Structured Generative Adversarial Networks. nov 2017.

[235] Ashish Vaswani, Google Brain, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need.

[236] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. LSTM Neural Networks for Language Modeling.

[237] Christopher J. Shallue, Jaehoon Lee, Joe Antognini, Jascha Sohl-Dickstein, Roy Frostig, and George E. Dahl. Measuring the Effects of Data Parallelism on Neural Network Training. nov 2018.

[238] Mike Schuster and Kuldip K Paliwal. Bidirectional Recurrent Neural Networks. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 45(11), 1997.

[239] Baoguang Shi, Xiang Bai, and Cong Yao. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. *arXiv Pre-print*, 8828(c):1–8, 2015.

[240] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-Chun Woo. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.

# Appendix A

# Publications

## Appendix B

# Experiment details for the Pixel-Wise Conditionned GAN

## Appendix C

# Experiment details for the Polarimetric CycleGAN