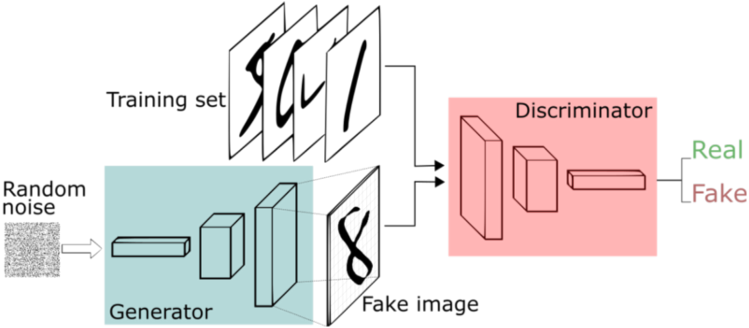
cDCGAN

conditional Deep Convolutional Generative Adversarial Networks

Edoardo De Matteis

Generative adversarial networks

The idea behind generative models is to learn a distribution from training data and generate new samples from it. GANs do it setting up a zero-sum game between a **discriminator** and a **generator**.


Training set
Discriminator
Real
Fake
Random noise
Generator
Fake image

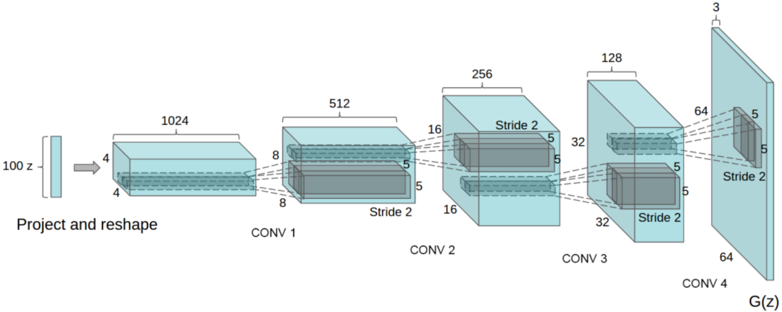
DCGAN

GANs where first introduced based only on FC layers, it came natural to extend them with convolutions when dealing with images.
To make them more stable to train we can constraint their architectural topology.

Architectural guidelines

- *All-convolutional net* i.e. replace deterministic spatial pooling with strided convolutions, allowing the network to learn its own spatial downsampling/upsampling.
- Eliminate fully connected layers on top of CNNs in favor of deeper architectures.

- Use of batch normalization in both the generator and discriminator, yet the first could be made unstable.
- Use ReLU activation in generator for all layers except for the output, which uses Tanh, and use LeakyReLU activation in the discriminator for all layers.



100 z
Project and reshape
1024
4
4
512
8
8
CONV 1
Stride 2
5
5
256
16
16
CONV 2
Stride 2
5
5
128
32
32
CONV 3
Stride 2
5
5
3
64
64
CONV 4
Stride 2
5
5
G(z)

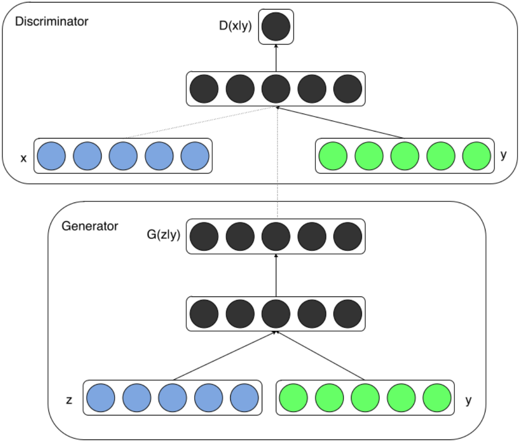
Results

It is hard in general to train GANs, after a few epochs with a small learning rate the generator learns well general features, yet it still far from realism.

LSUN

cGAN

One issue with GANs is that we cannot choose which output we get, hower by conditioning the model on additional information is possible to direct data generation.


Discriminator
D(x|y)
x
y
Generator
G(z|y)
z
y

cDCGAN

Combine the strengths of three models:

- *GAN*: Great generative model.
- *DCGAN*: Strong priors on images.
- *cGAN*: Conditional model.

Motivations

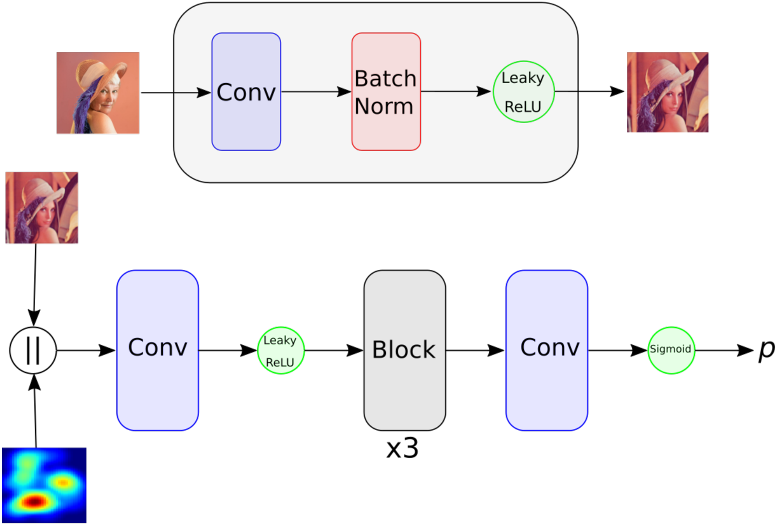
- Adding conditional data can have a regularizing effect.
- Specialized data enrichment.
- Entertainment e.g. *FaceApp*.
- Integration with other branches e.g. text to image.
- ...

Dataset

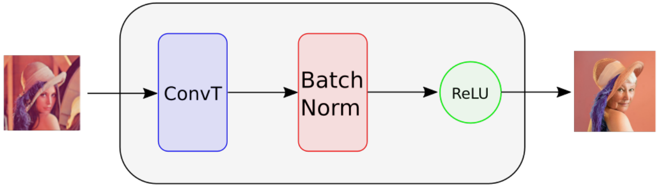
- MNIST.
- FashionMNIST.
- CIFAR10.

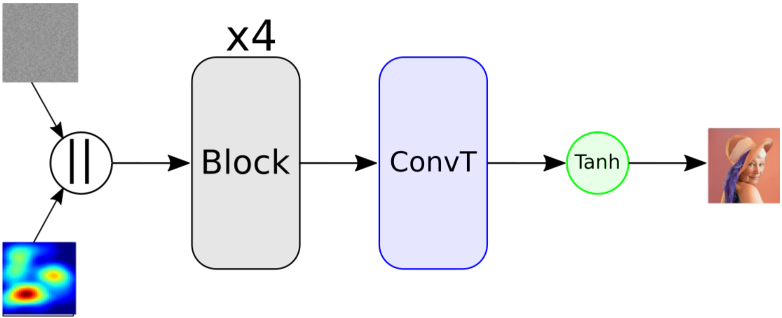
Data normalized by z-score and resized to 64×64 resolution. Random samples have dimension 64 and conditioning is directly concatenated as a channel to input vectors.

Discriminator



Generator





Loss

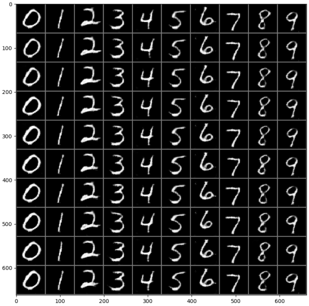
$$\mathcal{L}_{GAN} = \min_{\gamma} \max_{\delta} \mathbb{E}_{x \sim p_d} \log D(x) + \mathbb{E}_{z \sim \mathcal{N}} \log(1 - D(G(z))) \quad (1)$$

$$BCE = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(f(x_i)) + (1 - y_i) \log(1 - f(x_i)) \quad (2)$$

We define the ground truth on real and fake data as a vector of ones and zeros respectively.

MNIST

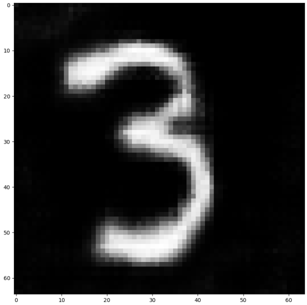
Weighting conditioning

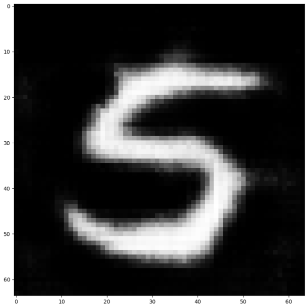


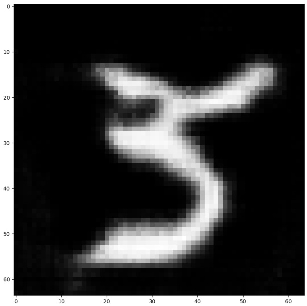
FashionMNIST

Doing the same...

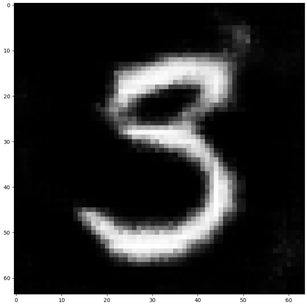
Multiclass conditioning

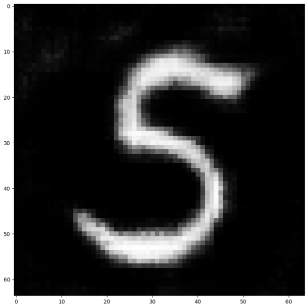



0
10
20
30
40
50
60
0 10 20 30 40 50 60

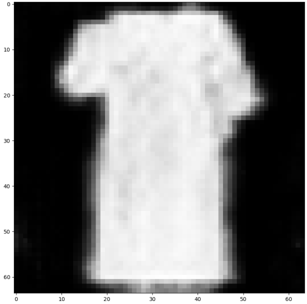

0
10
20
30
40
50
60
0
10
20
30
40
50
60

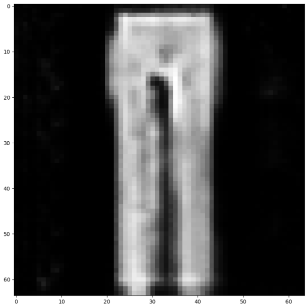
With different weights

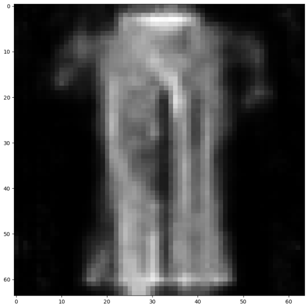



0
10
20
30
40
50
60
0
10
20
30
40
50
60

The same on FashionMNIST




0
10
20
30
40
50
60
0
10
20
30
40
50
60


0
10
20
30
40
50
60
0
10
20
30
40
50
60

Thank you

References

A. Radford, L. Metz, S. Chintala
Unsupervised representation learning with deep convolutional generative adversarial networks.

M. Mirza, S. Osindero
Conditional generative adversarial nets.