UNSUPERVISED REPRESENTATION LEARNING

WITH DEEP CONVOLUTIONAL

GENERATIVE ADVERSARIAL NETWORKS

ABSTRACT

1 INTRODUCTION

GANs provide an attractive alternative to maximum likelihood techniques.

In this paper, we make the following contributions:

We propose and evaluate a set of constraints on the architectural topology of Convolutional

GANs that make them stable to train in most settings. We name this class of architectures

Deep Convolutional GANs (DCGAN)

We use the trained discriminators for image classification tasks, showing competitive performance

with other unsupervised algorithms.

We visualize the filters learnt by GANs and empirically show that specific filters have

learned to draw specific objects.

We show that the generators have interesting vector arithmetic properties allowing for easy

manipulation of many semantic qualities of generated samples.

2 RELATED WORK

2.1 REPRESENTATION LEARNING FROM UNLABELED DATA

A classic approach to unsupervised representation learning is to do clustering on the data (for example using K-means), and leverage the clusters for improved classification scores.

Another popular method is to train auto-encoders (convolutionally, stacked, separating the what and where components of the code, ladder structures) that encode an image into a compact code, and decode the code to reconstruct the image as accurately as possible.

2.2 GENERATING NATURAL IMAGES

Generative image models are well studied and fall into two categories: parametric and nonparametric.

The non-parametric models often do matching from a database of existing images, often matching

patches of images.

Parametric models for generating images has been explored extensively.

2.3 VISUALIZING THE INTERNALS OF CNNS

One constant criticism of using neural networks has been that they are black-box methods, with little

understanding of what the networks do in the form of a simple human-consumable algorithm.

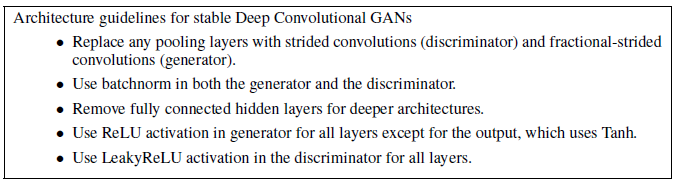
3 APPROACH AND MODEL ARCHITECTURE

Core to our approach is adopting and modifying three recently demonstrated changes to CNN architectures.

The first is the all convolutional net which replaces deterministic spatial pooling functions (such as maxpooling) with strided convolutions, allowing the network to learn its own spatial downsampling. We use this approach in our generator, allowing it to learn its own spatial upsampling, and discriminator.

Second is the trend towards eliminating fully connected layers on top of convolutional features.

Third is Batch Normalization which stabilizes learning by normalizing the input to each unit to have zero mean and unit variance.



4 DETAILS OF ADVERSARIAL TRAINING

We trained DCGANs on three datasets, Large-scale Scene Understanding, Imagenet-1k and a newly assembled Faces dataset.

5 EMPIRICAL VALIDATION OF DCGANS CAPABILITIES

5.1 CLASSIFYING CIFAR-10 USING GANS AS A FEATURE EXTRACTOR

One common technique for evaluating the quality of unsupervised representation learning algorithms

is to apply them as a feature extractor on supervised datasets and evaluate the performance of linear models fitted on top of these features.

5.2 CLASSIFYING SVHN DIGITS USING GANS AS A FEATURE EXTRACTOR

6 INVESTIGATING AND VISUALIZING THE INTERNALS OF THE NETWORKS

6.1 WALKING IN THE LATENT SPACE

The first experiment we did was to understand the landscape of the latent space.

6.2 VISUALIZING THE DISCRIMINATOR FEATURES

6.3 MANIPULATING THE GENERATOR REPRESENTATION

7 CONCLUSION AND FUTURE WORK

We think that extending this framework to other domains such as video (for frame prediction) and audio (pre-trained features for speech

synthesis) should be very interesting. Further investigations into the properties of the learnt latent

space would be interesting as well.