Presentation on "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," by Alec Radford, Luke Metz, and Soumith Chintala

John Hancock jhancoc4@fau.edu

November 27, 2018

Overview

- For the remainder of this presentation, we will refer to the paper entitled , "Unsupervised representational learning with deep convolutional generative adversarial networks," as, "the DCGAN's paper," or by its reference number [1], and Alec Radford, Luke Metz, and Soumith Chintala as, "the authors."

A generative adversarial network (GAN) is a neural network with two components. Goodfellow *et. al* invent GAN's in [5]. To a first approximation, GAN's work as follows:

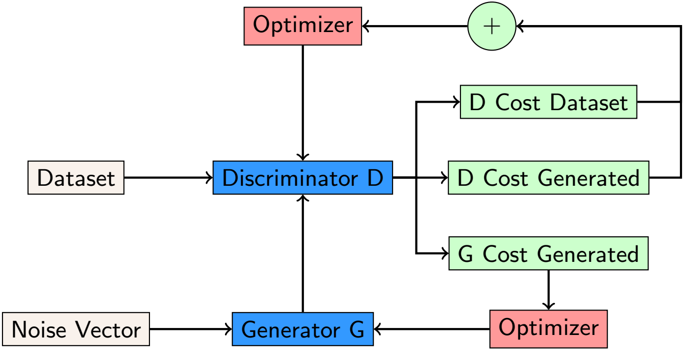
- The first component is a *generator* that learns to transform vectors of random numbers into output values that resemble instances from some dataset.
- The second component is a *discriminator* that classifies things into two categories:
  - the class of instances of the dataset, and
  - the class of generator outputs.
- "At convergence, the generators samples are indistinguishable from real data, and the discriminator outputs $\frac{1}{2}$ everywhere. The discriminator may then be discarded" [10]. from Deep learning , Goodfellow *et al.*

- ... or not. The authors of the DCGAN paper find a use for the discriminator.
- In the context of this paper, the outputs are images. However, researchers use GAN's where the generators create other artifacts. We find an extensive list on Github [11] of over 500 research projects. Some examples from this list are:
  - imputing missing values in datasets,
  - generating music,
  - fraud detection, and
  - playing chess.

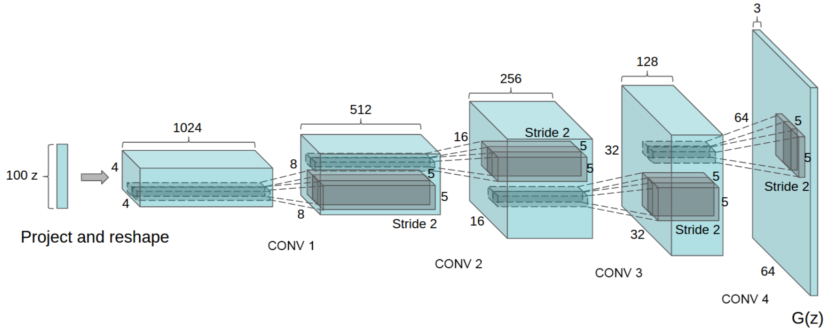
The authors of the paper make several contributions they...

- invent an architecture for DCGAN's,
- use the convolutional layer filters of trained DCGAN's discriminators as feature extractors for doing classifications,
- demonstrate that after training the DCGAN, its filters learn how to represent images, and
- present a method of doing vector arithmetic using DCGAN inputs to do inferences *à la* Word2Vec [6].

Architecture Diagram I



Generator Diagram from Paper [1]



Results I

- The authors use three datasets for training:
  - Large Scale Scene Understanding (LSUN),
  - Imagenet 1-K, and
  - Faces.
- The authors two datasets for evaluating unsupervised learning:
  Canadian Institute for Advanced Research (CIFAR) 10
  StreetView House Numbers (SVHN)

- The authors use images of bedrooms from the LSUN dataset [3] as input to their DCGAN.
- Then they find a way to identify and remove the generators feature maps [1] associated with windows. Note: see explanation of feature maps in Dr. Zhu's lectures on CNN's [12].
- After removing these feature maps, the generator no longer produces images of bedrooms with windows.
- The authors claim that this experiment proves the generator is learning representations of objects in an unsupervised manner.

- The authors did some processing known as guided backpropagation on the first 6 convolutional features of the last convolutional layer of their DCGAN's discriminator to create the image below.
- This is proof that the discriminator is learning features.
  - Note on the left-hand side taken before training that the pictures look like entire bedrooms, but after training the pictures have key portions of a bedroom scene, like windows and beds.
  - This demonstrates how the DCGAN is learning to represent objects in images.

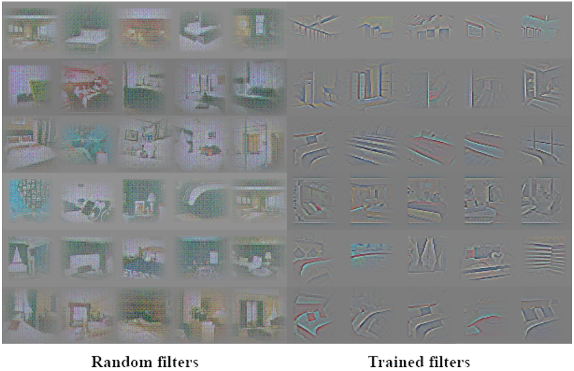


Figure: This image is coped from the paper [1].

Results - CIFAR I

- The authors use images from the **Imagenet 1-k** dataset as input to a DCGAN.
- After training, they take all the convolutional layers the discriminator learns, and use them as feature extractors.
- However they accomplished this, the authors then use the feature extractor they created from the DCGAN's discriminator for a support vector machine based classifier.
- They then used images from **CIFAR-10** dataset as inputs to this classifier that they report as 82.8% accurate. They mention similar K-Means based approaches that get lower accuracy so the result is competitive and noteworthy.
- This result is remarkable because they built a classifier that is able to correctly categorize images from one dataset, based on unsupervised learning methods involving a different dataset. It proves GAN's have generalization power.

Results - SVHN I

- We interpret the wording the authors use in section 5.2 to mean that the authors use images from the SVHN dataset as input to a DCGAN.
- It is not clear what the authors mean by, "non-extra set" they use for training their model.
- Is is our understanding that they train a DCGAN on the SVHN dataset, and then build a feature extractor similar to what they do with the DCGAN they build for Imagenet 1-k.
- Using this feature extractor, the authors then incorporate this into an L2-SVM classifier that we suppose is classifying images from SVHN.
- The notable result the authors report is that the classifier they build in this manner gets a lower error rate than a similar classifier built with a standard convolutional neural network, 22.48% vs. 28.87% respectively.

Results - Faces I

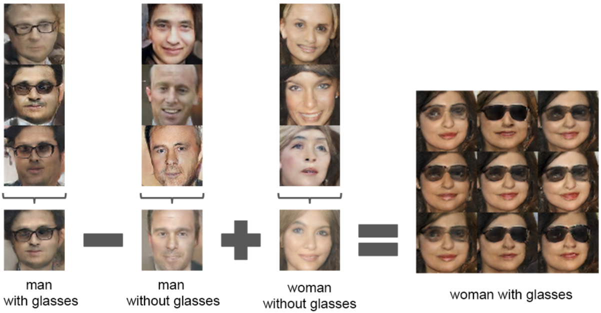
- The authors write that they create a faces dataset of images of faces from randomly selected web sites.
- After training the DCGAN on this dataset the authors do arithmetic on what they call, the "Z-vectors of sets of exemplar samples for visual concepts."
  - We take this to mean they were able to find groups of vectors of random numbers they used for inputs to the generator that produce images that look like something in particular, for example: a smiling man.
  - An important clue for our understanding of the term Z-vector is that the first input parameter of the generator function in the code that accompanies this paper [2] is named, "Z," and other vectors involving random numbers in the code also start with the letter, 'z.'

Results - Faces I

- The authors then do vector addition and subtraction with vectors they obtain from the average values of vectors in different exemplar sets.
- The authors then use the vectors are the results of these arithmetic operations as inputs to the generator.
- Please see the amazing result on the next slide. We feel this is the strongest result of the paper.

Results - Faces I

This image is from the DCGAN paper [1]:



Conclusions I

- The authors present a working architecture for GAN's that uses convolutional, and fractionally-strided (also known as deconvolutional) neural networks.
- Authors show how the convolutional and fractionally-strided convolutional layers indicate their DCGAN is automatically learning representations in an unsupervised manner.
- The source code that the authors wrote is publicly available and it bolsters their claims; one can access the source code on Github [2], or a derivative work such as [4] and confirm that it generates outputs as advertised.

References I

S. Chintala, L. Metz, and A. Radford, Unsupervised representation learning with deep convolutional generative adversarial networks. 2016. [Online]. Available: arXiv:1511.06434v2 [cs.LG]

S. Chintala, L. Metz, and A. Radford, dcgan_code. (2016, May). Available: https://github.com/Newmu/dcgan_code. [Accessed Nov. 19, 2018].

Large Scale Scene Understanding Challenge (2017, July). Available: http://lsun.cs.princeton.edu/2017/. [Accessed Nov. 19, 2018].

T Kim. (2018, August). Available: https://github.com/carpedm20/DCGAN-tensorflow. [Accessed Nov. 19, 2018].

Y. Bengio, A. Courville,I. Goodfellow, M. Mirza, S. Ozair, J. Pouget-Abadie, D. Warde-Farley, B. Xu, Generative adversarial nets. 2014. [Online]. Available: arXiv:1406.2661v1 [stat.ML]

References II

- T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Advances in Neural Information Processing Systems 26 (NIPS 2013), 2013. [Online] Available: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf
- Creodocs Limited, "Beamer Presentation," *latextemplates.com*, 2018. [Online], Available: http://www.latextemplates.com/templates/presentations/1/presentation_1.zip. [Accessed Nov. 10, 2018].
- IEEE, Piscataway, NJ, USA. *IEEE Editorial Style Manual*. 2016. [Online]. Available: http://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE_Style_Manual.pdf, [Accessed Nov. 11, 2018].
- F. Chollet, and J.J. Allaire. *Deep Learning with R*. Manning Publications 2018. [E-Book] Available: Safari E-Book.

References III

Y. Bengio, A. Courville, I. Goodfellow. (2016). "Chapter 20 Deep Generative Models," 2016. [Online] Available: https://www.deeplearningbook.org/contents/generative_models.html. [Accessed: Nov. 20, 2018].

A. Hindupur, "A list of all named GANs!" github.com, Sep. 30, 2018. [Online]. Available: https://github.com/hindupuravinash/the-gan-zoo. [Accessed Nov. 21, 2018].

Xingquan Zhu. 2018. Convolutional Neural Networks (CNN) retrieved October 22, 2018 from https://canvas.fau.edu/files/14310707/download?download_frd=1

The End