Week10 생성모델 과제 – 논문 리뷰

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

Tobigs 14기 이혜린

▶ Abstract

- Supervised CNN에 비해 unsupervised CNN은 관심 받지 못했다. 이 차이를 줄이기 위해 GAN을 소개한다.

▶ 1. Introduction

- GAN을 학습시키고 generator와 discriminator의 부분들을 supervised tasks를 위한 feature extractor로 후에 재사용할 수 있는 image representation 방법을 소개할 것이다.
- GAN은 train이 불안정해서 generator의 output이 종종 nonsensical 하다는 단점이 있다.

▶ 2. Related Work

- Representation learning from unlabled data
- Generating natural images
- Visualizing the internals of CNNs

▶ 3. Approaching and Model Architecture

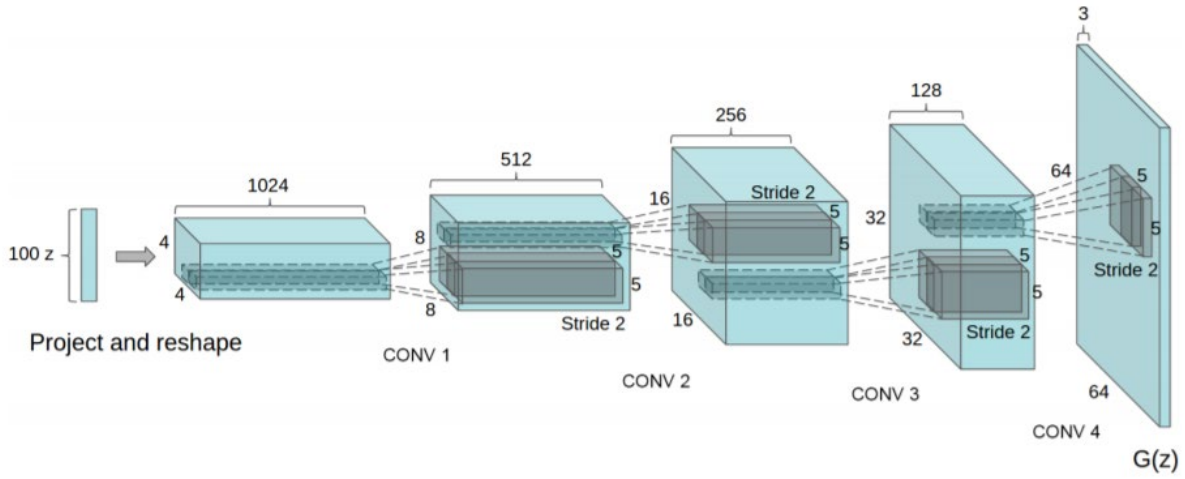

Figure 1: DCGAN generator used for LSUN scene modeling. A 100 dimensional uniform distribution Z is projected to a small spatial extent convolutional representation with many feature maps. A series of four fractionally-strided convolutions (in some recent papers, these are wrongly called deconvolutions) then convert this high level representation into a 64×64 pixel image. Notably, no fully connected or pooling layers are used.

- DCGAN은 기존의 GAN과 다음과 같은 차별점을 만들었다.
  1. Pooling layer를 사용하지 않았다.
  2. Fully connected layer를 사용하지 않았다. (input z가 이미 flatten layer 형태이기 때문)

  → generator에는 fractional-strided, discriminator에는 strided convolutional layer 사용

  3. Batch Normalization를 사용했다. (Instability를 방지하기 위해 generator output layer와 discriminator input layer에는 적용하지 않았다.)

  → poor initialization과 gradient flow로 인해 생기는 training problem, instability 해결.

- Activation
  1. Generator : ReLU (마지막 output layer만 tanh)
  2. Discriminator : LeakyReLU (alpha=0.2)

▶ 4. Details of Adversarial Training

- 3개의 데이터셋으로 train. (LSUN, Imagenet-1k, and a newly assembled Faces dataset)
- Mini_batch SGD : batch size : 128
- 모든 가중치는 평균 0, 표준편차 0.02인 정규분포에 의해 초기화
- Adam Optimizer (lr = 0.0002, momentum=0.5)

▷ 4.1 LSUN

- 300만개의 침실 데이터셋
- 모델의 속도와 일반화 peroformance는 직접적인 관계가 있다.
- No data augmentation
- 모델이 과적합/memorizing training으로 인한 높은 퀄리티의 샘플을 생산하지 않는 것을 보여주기 위해 샘플을 보여줄 것이다.
- Deduplication : 단순 암기를 통해 샘플을 생성하는 것을 방지하기 위한 단계이다. Autoencoder로 image를 코드로 표현한 후 중복되는 것은 삭제. 총 27만 5천개를 삭제했고 precision이 0.01 상승했다.

Figure 2: Generated bedrooms after one training pass through the dataset. Theoretically, the model could learn to memorize training examples, but this is experimentally unlikely as we train with a small learning rate and minibatch SGD. We are aware of no prior empirical evidence demonstrating memorization with SGD and a small learning rate.

Figure 3: Generated bedrooms after five epochs of training. There appears to be evidence of visual under-fitting via repeated noise textures across multiple samples such as the base boards of some of the beds.

▷ 4.2 Faces & 4.3 Imagenet-1k

▶ 5. Empirical Validation of DCGANs Capabilities

▷ 5.1 Classifying CIFAR-10 using GANs as a feature extractor

- 비지도 representation learning을 평가하는 방법 중 하나는 알고리즘을 supervised datasets에 feature extractor로 적용하고 이 특성들의 linear model의 모델을 평가하는 것.
- CIFAR-10이 아닌 Imagenet-1k로 훈련되었고, features는 CIFAR-10을 분류하는 데에 사용되었다.
- Exemplar CNN 보다 성능이 낮았으나, 이는 후에 보완해야 할 부분으로 생각된다.

▷ 5.2 Classifying SVHN digits using GANs as a feature extractor

- DCGAN의 discriminator의 특성들을 사용했다.
- Test Error rate 22.48. (다른 CNN 보다 뛰어남)

▶ 6. Investigating and Visualizing the Internals of the Networks

▷ 6.1 Walking in the Latent Space

- Latent space를 순차적으로 관찰
- 물체가 갑자기 추가되거나 사라지면, 계층적으로 무너졌다고 볼 수 있다. 즉, 단순 암기로 인한 샘플 생성이 이루어지고 있다고 볼 수 있다.

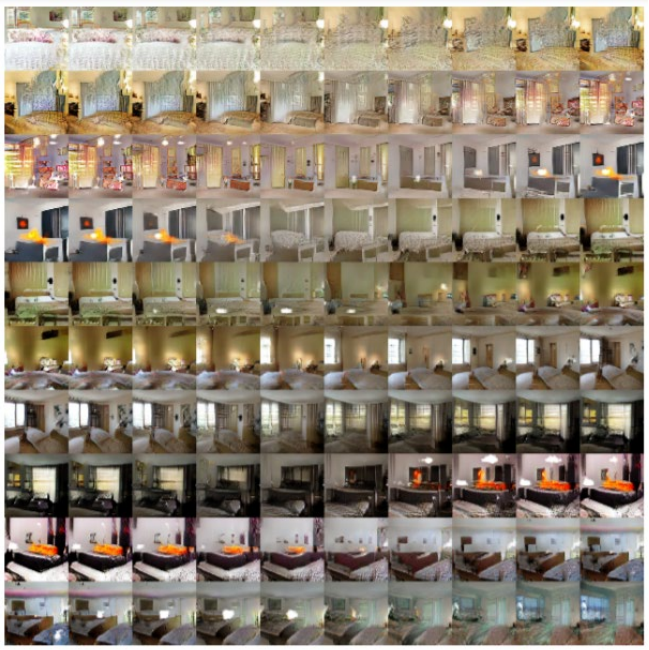

Figure 4: Top rows: Interpolation between a series of 9 random points in Z show that the space learned has smooth transitions, with every image in the space plausibly looking like a bedroom. In the 6th row, you see a room without a window slowly transforming into a room with a giant window. In the 10th row, you see what appears to be a TV slowly being transformed into a window.

▷ 6.2 Visualizing the Discriminator features

- Backpropagation을 통해, discriminator가 중점적으로 보고 있는 부분이 어디인지를 볼 수 있다.

Figure 5: On the right, guided backpropagation visualizations of maximal axis-aligned responses for the first 6 learned convolutional features from the last convolution layer in the discriminator. Notice a significant minority of features respond to beds - the central object in the LSUN bedrooms dataset. On the left is a random filter baseline. Comparing to the previous responses there is little to no discrimination and random structure.

▷ 6.3 Manipulating the Generator Representation

◇ 6.3.1 Forgetting to Draw Certain Objects

- Generator가 window를 이미지에서 잘 제거하는지 실험. 어떤 sample은 창문이 잘 지워졌으나, 어떤 샘플은 창문은 지워졌지만 창문 대신 다른 물체로 변경됨
- 이미지 품질은 저하되었으나, 이미지의 전체적인 장면이 유지되는 것으로 보아 generator가 잘 작동하고 있다는 것을 알 수 있음

Figure 6: Top row: un-modified samples from model. Bottom row: the same samples generated with dropping out "window" filters. Some windows are removed, others are transformed into objects with similar visual appearance such as doors and mirrors. Although visual quality decreased, overall scene composition stayed similar, suggesting the generator has done a good job disentangling scene representation from object representation. Extended experiments could be done to remove other objects from the image and modify the objects the generator draws.

◇ 6.3.2 Vector Arithmetic on Face Samples

- Generator의 Z representation으로 이미지의 벡터 연산이 가능한 것을 증명함. Z를 평균내서 하는게 더 안정적이었다.

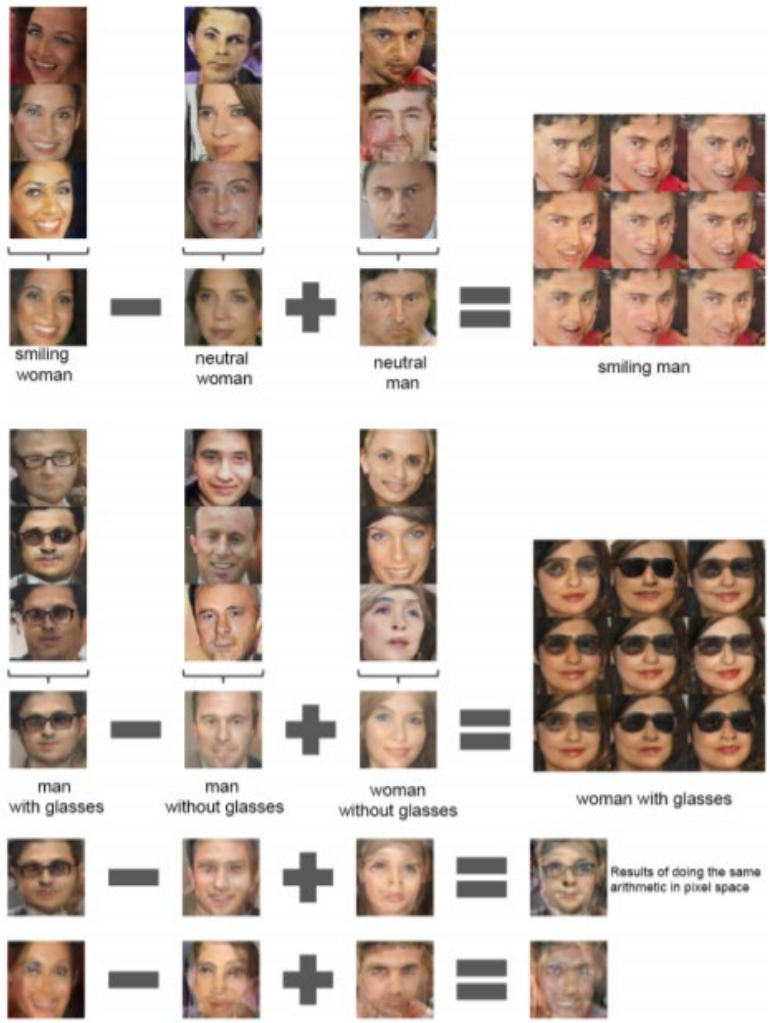

Figure 7: Vector arithmetic for visual concepts. For each column, the Z vectors of samples are averaged. Arithmetic was then performed on the mean vectors creating a new vector Y. The center sample on the right hand side is produce by feeding Y as input to the generator. To demonstrate the interpolation capabilities of the generator, uniform noise sampled with scale +-0.25 was added to Y to produce the 8 other samples. Applying arithmetic in the input space (bottom two examples) results in noisy overlap due to misalignment.

▶ 7. Conclusion and Future Work

- 위 논문은 GAN을 안정적으로 훈련시킬 수 있는 네트워크를 소개했으며, 적대 신경망이 supervised learning과 generative modeling에서 이미지의 좋은 표현들을 학습한다는 것을 확인했다.
- 그러나 여전히 불안정성이라는 문제가 남아있다. 모델이 더 오래 훈련되면 언젠가는 필터의 일부분이 진동하는 형태로 무너질 수도 있다.
- 이것을 해결하기 위해서, 이 프레임워크를 비디오, 오디오 등의 다른 분야들로 확장해보는 것도 좋을 것이라고 생각한다.