

UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS

Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434* (2015).

ABSTRACT

cnn을 활용하여 deep convolutional gan 구축

1. INTRODUCTION

비지도 학습에서 unlabeled dataset의 재사용 가능한 특징 추출이 필요함.

즉, unlabeled image를 활용하여 특징 표현을 한다면, 이미지 분류 등의 지도학습에서 사용될 수 있다.

본 논문에서는 이를 위해 gan을 활용하여 image representation을 학습 하는 방법을 제안하였다.

- 이미지 분류에 학습된 분류기를 사용하고, 비지도 학습과 비교하여 경쟁력 있는 성능이 나올지 보여줌.
- gan이 학습된 필터를 시각화하고, 특정 필터가 특정 객체를 그리도록 학습함을 보여줄 수 있음.
- generator가 생성된 샘플의 특성들은 산술연산을 가져 의미적 특성들의 조작이 쉬움.

3. APPROACH AND MODEL ARCHITECTURE

본 논문에서의 3가지 중요한 접근방식은 아래와 같다.

1. maxpooling 등의 기능을 stride convolution을 사용하여 down sampling을 학습한다.
2. convolutional feature 위의 fully connected layer를 제거한다.
3. batch normalization을 사용한다. batch normalization은 각 유닛의 입력을 정규화 하여 각 층의 잘못된 초기화로 발생하는 문제와, 더 많은 층을 가진 모델에서 기울기를 전파하는데 도움을 주어 학습을 안정화시킬 수 있다. 하지만, gan에서 모든 층에 적용한다면, sample oscillation과 model instability의 문제가 발생하여, generator의 output과 discriminator의 input엔 사용하지 않는다.

generator에서는 relu를 사용하고 마지막 layer만 tanh를 사용하였다. 이는 generator가 색상 공간에 있어 빨리 학습하는데 도움이 되었다.

discriminiator에서는 leaky relu를 사용하였고, 마지막 layer에서는 sigmoid로 사용하였다.

4. DETAILS OF ADVERSARIAL TRAINING

drgan을 학습하는데 있어 논문에서는 image를 generator의 output의 범위인 [-1,1]로 scaling을 진행하였으며, 128의 mini batch, sgd(학습률 0.0002)를 사용했다고 한다. 또한 momentum term을 0.9가 아닌 0.5를 사용하는 것이 학습의 안정성에 있어서 도움이 되었다고 한다.

4.1 LSUN (침실 이미지 데이터셋)

이미지 품질이 향상되면서 overfitting/memorizing training examples 문제가 우려됨



[fig 2]

위의 그림은 1에폭을 통해 생성한 영상이지만, 적은 학습률을 사용하였기 때문에 memorizing의 문제가 아니다.



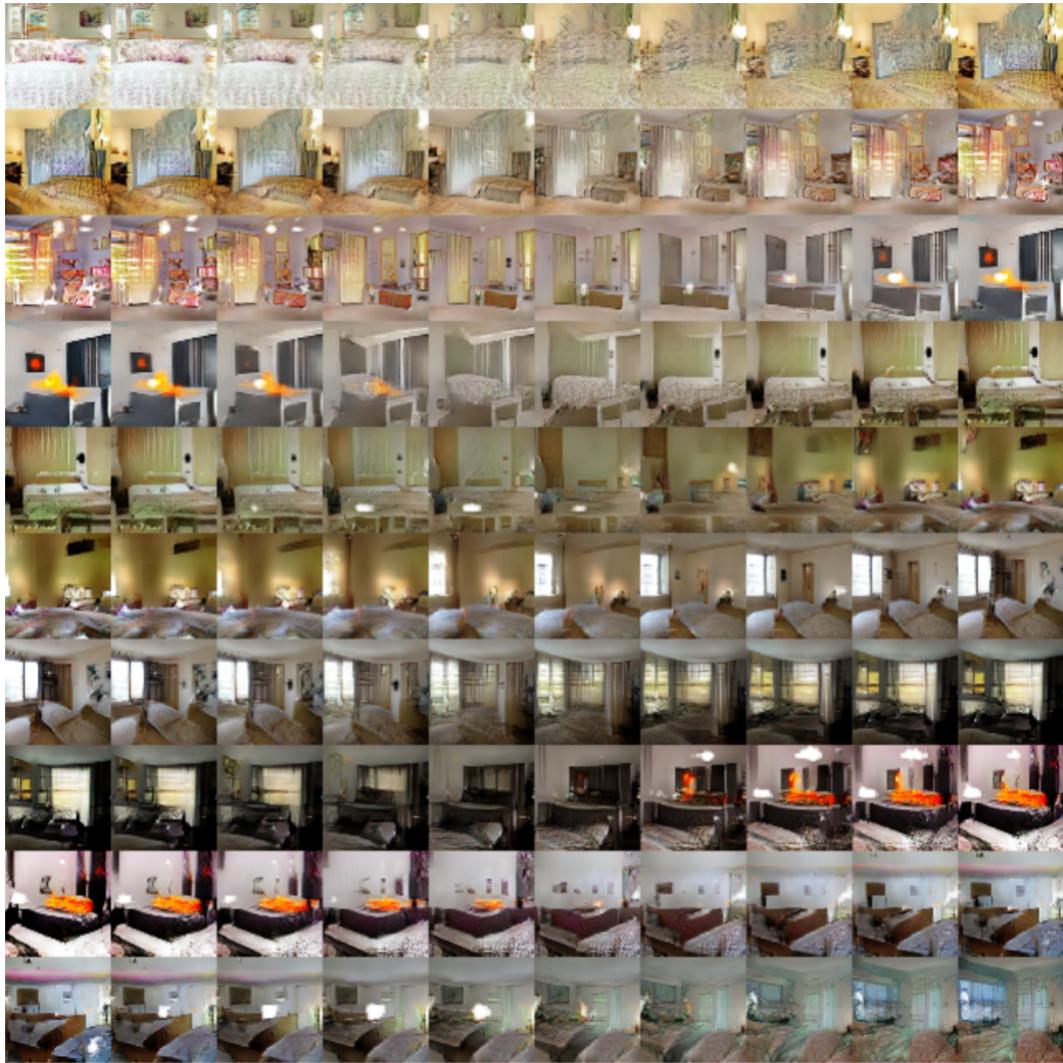
[fig 3]

위 그림은 5에폭을 학습한 결과로 몇몇 침대 위의 노이즈로 under-fitting으로 판단 되어짐.

INVESTIGATING AND VISUALIZING THE INTERNALS OF THE NETWORKS

6.1 WALKING IN THE LATENT SPACE

본 논문에서의 실험은 벡터 Z 의 값을 조금씩 바꿔갔을 때 생성되는 이미지의 결과가 부드럽게 변경되어짐을 보여줍니다.



[fig 4]

위의 그림처럼 서서히 변해감을 확인할 수 있고, 가장 아래의 이미지에서는 tv가 창문으로 서서히 변해감을 확인할 수 있어, 모델은 memorization이 없다고 확인할 수 있다.

6.2 VISUALIZING THE DISCRIMINATOR FEATURES



[fig 5]

Random filters

Trained filters

위의 그림은 dcgan이 학습 시에 무작위의 필터를 학습한 것이 아닌, 침대, 창문 등의 계층적 구조를 학습한 것을 보여주고 있다.

MANIPULATING THE GENERATOR REPRESENTATION

FORGETTING TO DRAW CERTAIN OBJECTS

generator에서의 특정 객체 표현을 학습한 형태를 알아보기 위해 generator에서 특정 객체를 나타내는 필터를 제거(dropout 사용)하는 실험을 수행하였다.

실험 방법은,

1. 150개의 sample image에서 52개의 창문을의 위치를 지정하는 bounding box를 만든다.
2. bounding box를 고차원의 특징을 갖는 conv layer를 통과 시켰을 때 positive(양수)의 결과를 갖는 filter를 추출한다.
3. 추출한 filter를 dropout 시킨다.



[fig 6]

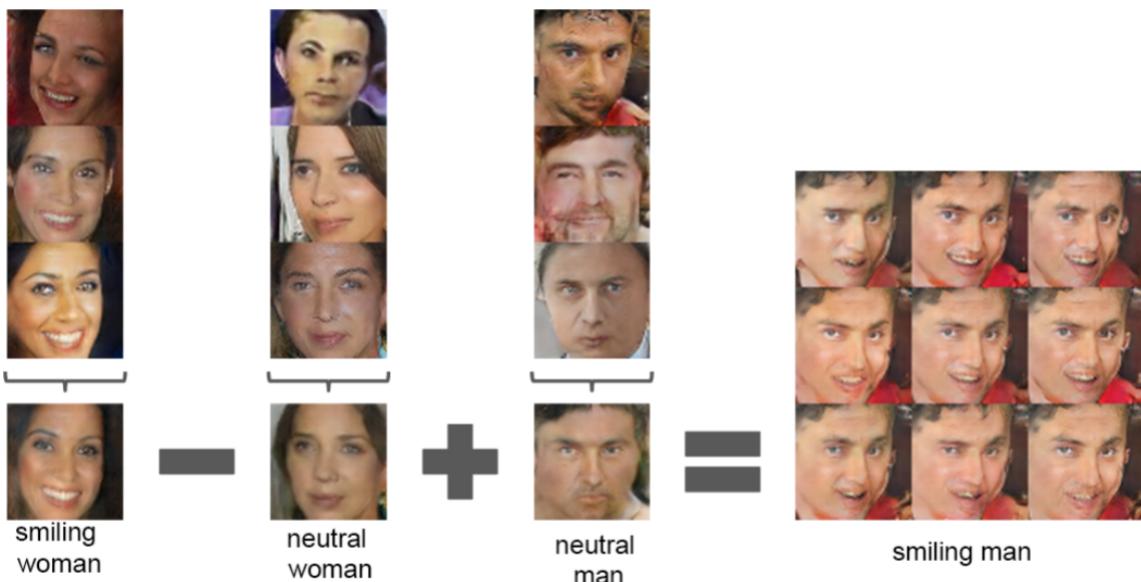
위의 그림은 창문에 해당하는 filter를 삭제시킨 결과로 실제로 창문의 형태가 거의 사라짐을 확인할 수 있다. 이러한 실험은 생성된 이미지에서 특정 객체를 수정할 수 있다는 것을 보여줄 수 있다.

VECTOR ARITHMETIC ON FACE SAMPLES

word2vec에서의 벡터의 예시로 아래와 같이 왕이지만, 남자는 아니고 여자인 벡터는 여왕이다. 라는 예시가 있는데 이것이 dcgan에서 잠재벡터 Z가 유사한 결과를 도출하는지에 대한 실험을 진행하였다.

$$\begin{aligned} \text{vector}(King) - \text{vector}(Man) + \text{vector}(Woman) \\ = \text{vector}(Queen) \end{aligned}$$

실험은 웃는 여자의 평균 벡터 - 일반 여자의 평균 벡터 + 일반 남자의 평균 벡터이다.



[fig 7]

이러한 실험 결과와 같이 Z에 대해서 벡터의 산술 연산을 통해 특징의 변환을 할 수 있음을 알 수 있다.

CONCLUSION

이 논문은 gan이 mode collapse()등의 불안정성을 가지지만, 이미지 생성을 위해서 원본 이미지의 좋은 특징(표현)을 학습한다는 증거를 제공하였다.