INFOGAN

DISENTANGLED REPRESENTATIONS

REFERENCE

- Chen, Xi, et al.
  "Infogan: Interpretable representation learning by information maximizing generative adversarial nets."
  Advances in neural information processing systems.
  2016.

INTRODUCTION

INTRODUCTION

- CGAN, ACGAN
  - 생성기 출력 제어
  - 원하는 숫자를 생성할 수 있음
- 출력의 특성(features)을 제어할 수는 없음

INTRODUCTION

- 출력의 특성을 지정할 수 있는 GAN
  - 분해된 표현(disentangled representations) GAN
- InfoGAN
  - Interpretable representation
  - learning by Information maximizing
  - generative adversarial nets

DISENTANGLED REPRESENTATIONS

DISENTANGLED REPRESENTATIONS

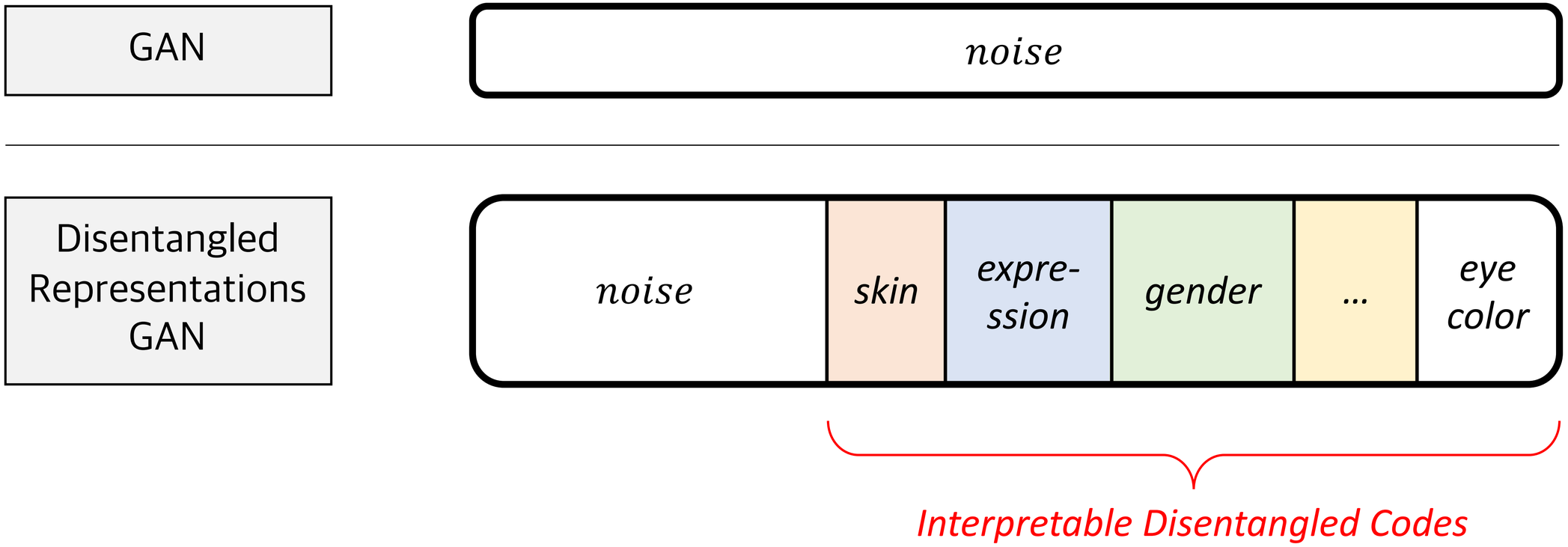
- 사람 얼굴의 여러 특성
  - 피부 색
  - 눈동자 색
  - 표정
  - 등

DISENTANGLED REPRESENTATIONS

- 지금까지 공부한 GAN으로는 특성을 지정할 수 없음
  - 특성을 생성기에 요청할 수 없음
  - 노이즈를 균등분포에서 임의로 샘플링해서 이미지 생성
- 노이즈에 해당하는 얽힌 잠재 코드를 분해된 코드로
  - 코드: 의미를 지닌 벡터로 취급

DISENTANGLED REPRESENTATIONS

- 얽힌 코드와 분해된 코드



DISENTANGLED REPRESENTATIONS

- 조건을 포함한 생성기 출력 $G(z, c)$ 대신:
  - $G(\mathbf{z})$ where $\mathbf{z} = (z, c)$, $c = c_1, c_2, \ldots, c_L$
- 모든 잠재 코드들 $c_1, c_2, \ldots, c_L$은 독립적이라 가정
  - 문제를 단순화
  - $p(c_1, c_2, \cdots, c_L) = \prod_{i=1}^{L} p(c_i)$

MUTUAL INFORMATION

MUTUAL INFORMATION

- 정보 이론
- 상호 정보량(Mutual Information, MI)
  - 엔트로피 H에 대하여
  - $I(X;Y) = H(X) - H(X|Y)$
  - Y가 관측됐을 때 X로부터 사라지는 불확실성(엔트로피)의 양

MUTUAL INFORMATION

- InfoGAN에서 상호 정보량이 최대화되도록 함
  - $I(c; G(z, c)) = H(c) - H(c \mid G(z, c))$
- (=) $H(c \mid G(z, c))$가 작은 값을 가지도록 함
- (=) 생성된 출력 G(~) 관측에 빠른 잠재 코드(c)의 불확실성을 줄임
  - 원하는 바, 생각하는 바를 잘 만들어 냄

MUTUAL INFORMATION

- GAN과 InfoGAN의 손실 함수
- GAN
  - $L^{(D)} = -\mathbb{E}_{x \sim P_{data}} log D(x) - \mathbb{E}_z log(1 - D(G(z)))$
  - $L^{(G)} = -\mathbb{E}_z log D(G(z))$

MUTUAL INFORMATION

- GAN과 InfoGAN의 손실 함수
- InfoGAN
  - $L^{(D)} = -\mathbb{E}_{x \sim P_{data}} log D(x) - \mathbb{E}_{z,c} log(1 - D(G(z,c))) - \lambda I(c; G(z,c))$
  - $L^{(G)} = -\mathbb{E}_{z,c} log D(G(z,c)) - \lambda I(c; G(z,c))$

INFOGAN

INFOGAN

- 문제는 $\lambda I(c; G(z, c))$ 계산이 어렵다는 것
  - 상호정보량 계산에
  - 사후 분포 $P(c \mid G(z, c)) = P(c \mid x)$의 계산이 필요
  - 모르는 정보
    - 일반적으로 $p(x)$ 계산이 불가하기 때문

INFOGAN

- 상호 정보량을 직접 계산하기는 불가능하니
  - 하한(lower bound)를 계산하고
  - 하한을 최대화하는 방식을 사용

INFOGAN

- $P(c \mid x)$ 대신 $Q(c \mid x)$ 사용
  - 하한
  - 보조적 분포(auxiliary distribution)
  - $P(c \mid x)$를 근사

INFOGAN

- $I(c; G(z, c)) = H(c) - H(c \mid G(z, c))$
  - 전개 후 정리하면
  - 생략
- $I(c; G(z, c)) \geq L_I(G, Q) = E_{c \sim P(c), x \sim G(z,c)}[logQ(c \mid x)] + H(c)$

INFOGAN

- $L_I(G, Q) = E_{c \sim P(c), x \sim G(z,c)}[logQ(c \mid x)] + H(c)$
- $L_I(G, Q)$는 $P(c \mid x)$와 관련 없음
  - 계산 가능
- 하한 최댓값은 $H(c)$
- G에 대해서 Q를 최대화하는 문제로 바뀜
  - GAN 학습에 포함되기 자연스러움

INFOGAN

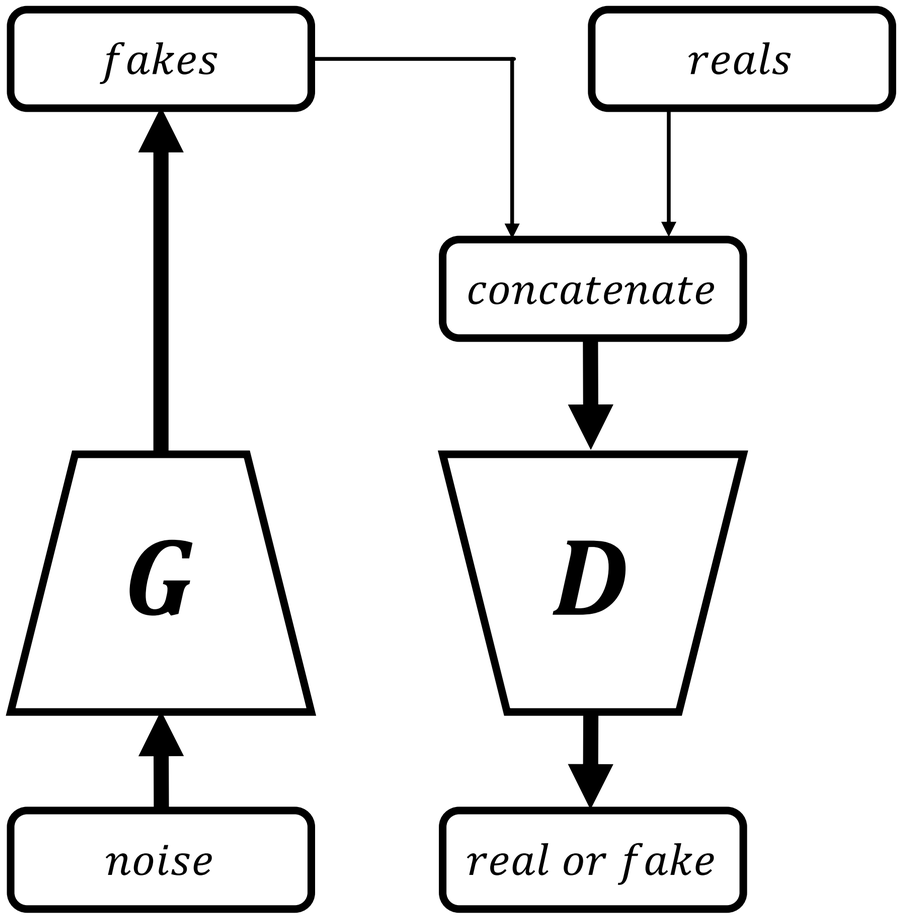
- Q를 어떻게 구현할 것인가?
  - 보조 신경망을 활용
  - $Q(c \mid x; \theta)$

INFOGAN

- 보조 신경망을 마지막 계층에 추가
  - 계산량 증가는 미미
  - 훈련 시간 등에 영향을 거의 끼치지 않음
  - 매우 빨리 수렴함

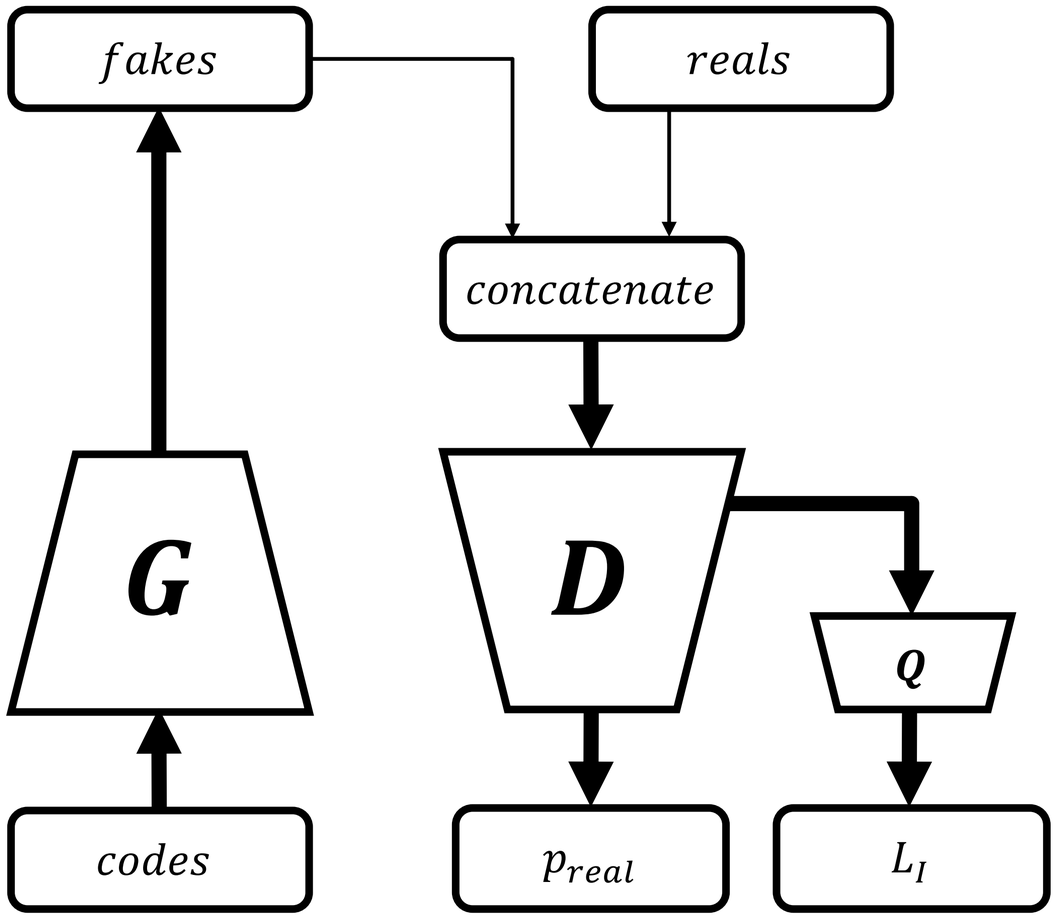
INFOGAN

- GAN 다이어그램:



INFOGAN

- InfoGAN 다이어그램:



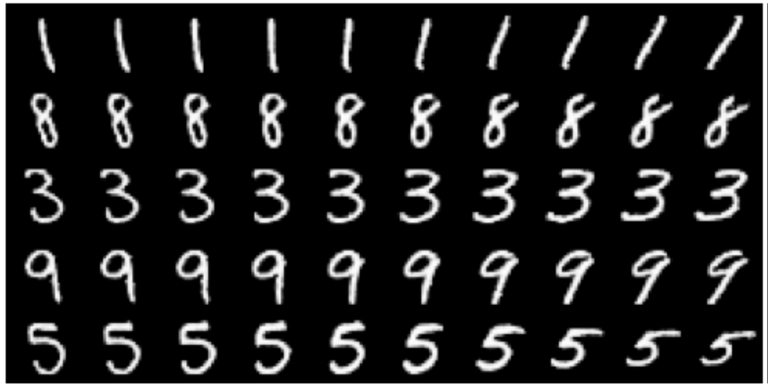
EXPERIMENTS

EXPERIMENTS

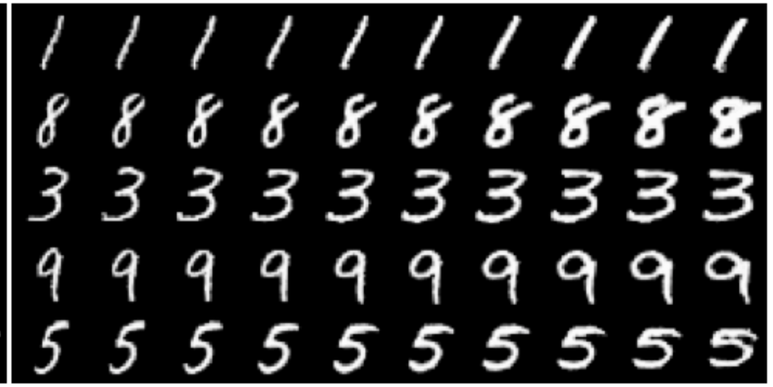
- MNIST에서 잠재 코드 조정:

(a) Varying c_1 on InfoGAN (Digit type)

(b) Varying c_1 on regular GAN (No clear meaning)



(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)



(d) Varying c_3 from -2 to 2 on InfoGAN (Width)

EXPERIMENTS

- 3D 얼굴에서 잠재 코드 조정:

(a) Azimuth (pose)

(c) Lighting

(b) Elevation

(d) Wide or Narrow

EXPERIMENTS

- 3D 의자에서 잠재 코드 조정:
  - 폭을 조정하니 자연스럽게 소파—>의자

(a) Rotation (b) Width

EXPERIMENTS

- CelebA에서 잠재 코드 조정:

(a) Azimuth (pose)

(b) Presence or absence of glasses

(c) Hair style

(d) Emotion

CONCLUSION

CONCLUSION

- InfoGAN은
  - 비지도적이고 해석가능한 학습을 수행
  - GAN에서 아주 작은 정도의 연산만을 추가로 요구
- VAE 등에서도 활용 가능

CONCLUSION

- 계층별 잠재 표현을 학습하거나
- 더 나은 코드를 활용한 준(semi)-지도 학습,
- 고차원 데이터 디스커버리 툴 등으로도 활용 가능할 것

INFOGAN

DISENTANGLED REPRESENTATIONS