WATCH YOUR UP-CONV

CNN BASED GENERATORS ARE DOOMED

BEFORE READING PAPER

- Ref: Durall, Ricard, Margret Keuper, and Janis Keuper. "Watch your Up-Convolution: CNN Based Generative Deep Neural Networks are Failing to Reproduce Spectral Distributions." arXiv preprint arXiv:2003.01826 (2020).
- 푸리에 변환 ref: https://darkpgmr.tistory.com/171

BACKGROUND

- 이미지
  - 공간 영역(Spatial domain): 일반적으로 생각하는 이미지
  - 변환 영역(Transform domain): 주파수 영역

BACKGROUND

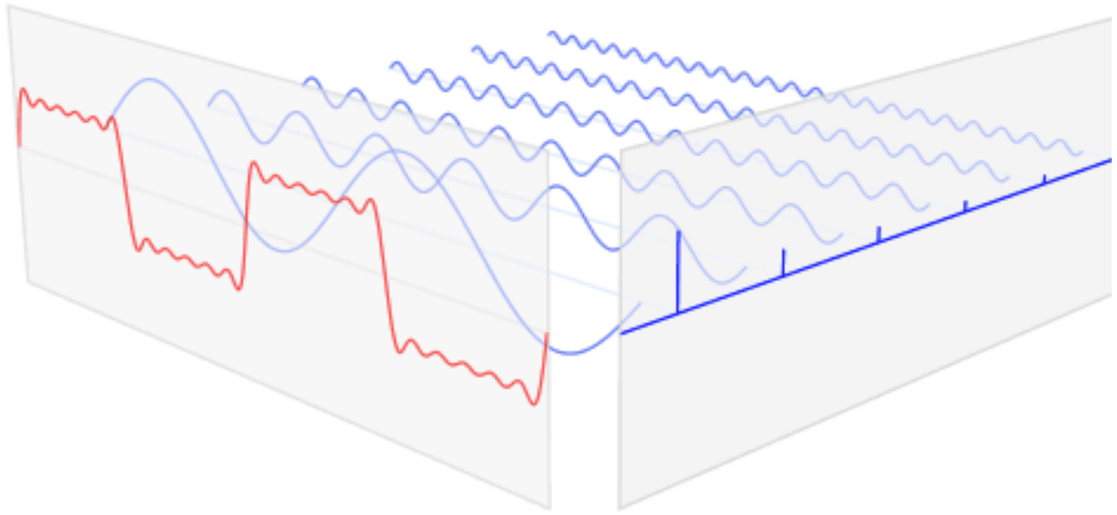
- 공간 주파수(Spatial frequency)
  - 공간 영역 상에서 화소 밝기의 변화율
  - 고주파 영역: 밝기가 빨리 변화하는 영역 - 주로 경계 부분
  - 저주파 영역: 밝기가 거의 변하지 않는 영역

BACKGROUND

- 파워 스펙트럼(power spectrum)
  - 스펙트럼이라 부르는 시계열 분석의 정확한 명칭

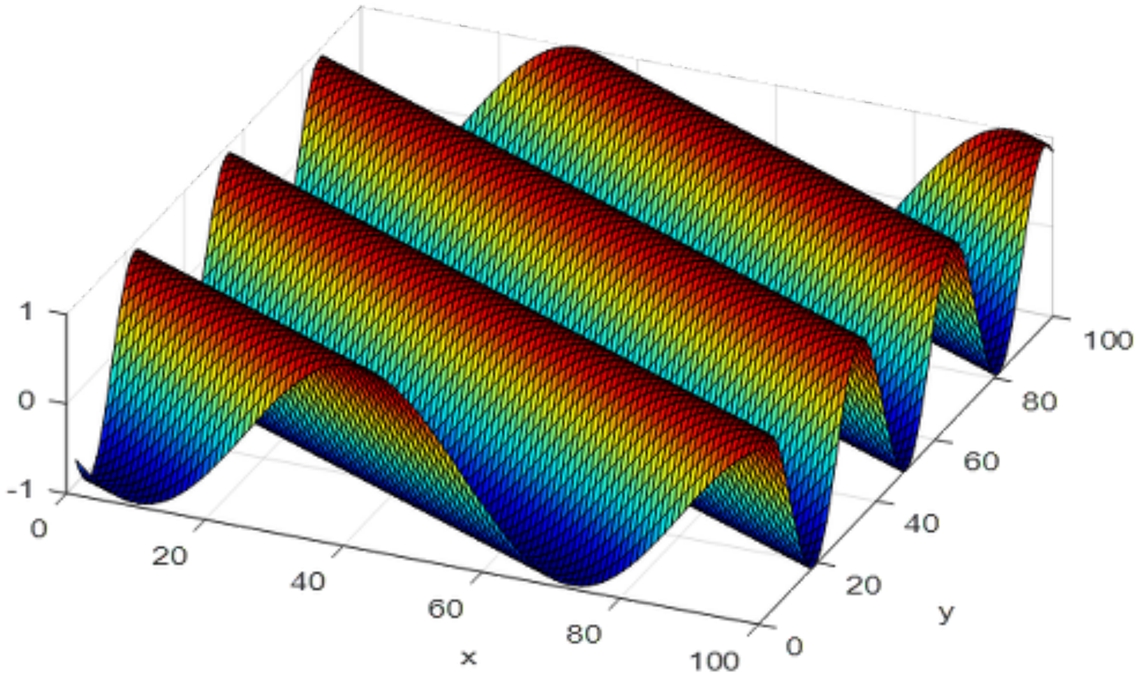
BACKGROUND

- 푸리에 변환(Fourier transform): 푸리에 급수에서 주기 T를 무한으로 생각(비주기 함수 격)
  - 임의의 입력 신호를 다양한 주파수를 가지는 주기함수들의 합으로 분해
    - 주기함수: sin, cos 삼각함수
  - 영상을 주파수 성분으로 변환



BACKGROUND

- 이미지에 대한 푸리에 변환
  - 다양한 2D 정현파(sinusoidal wave)들의 합



BACKGROUND

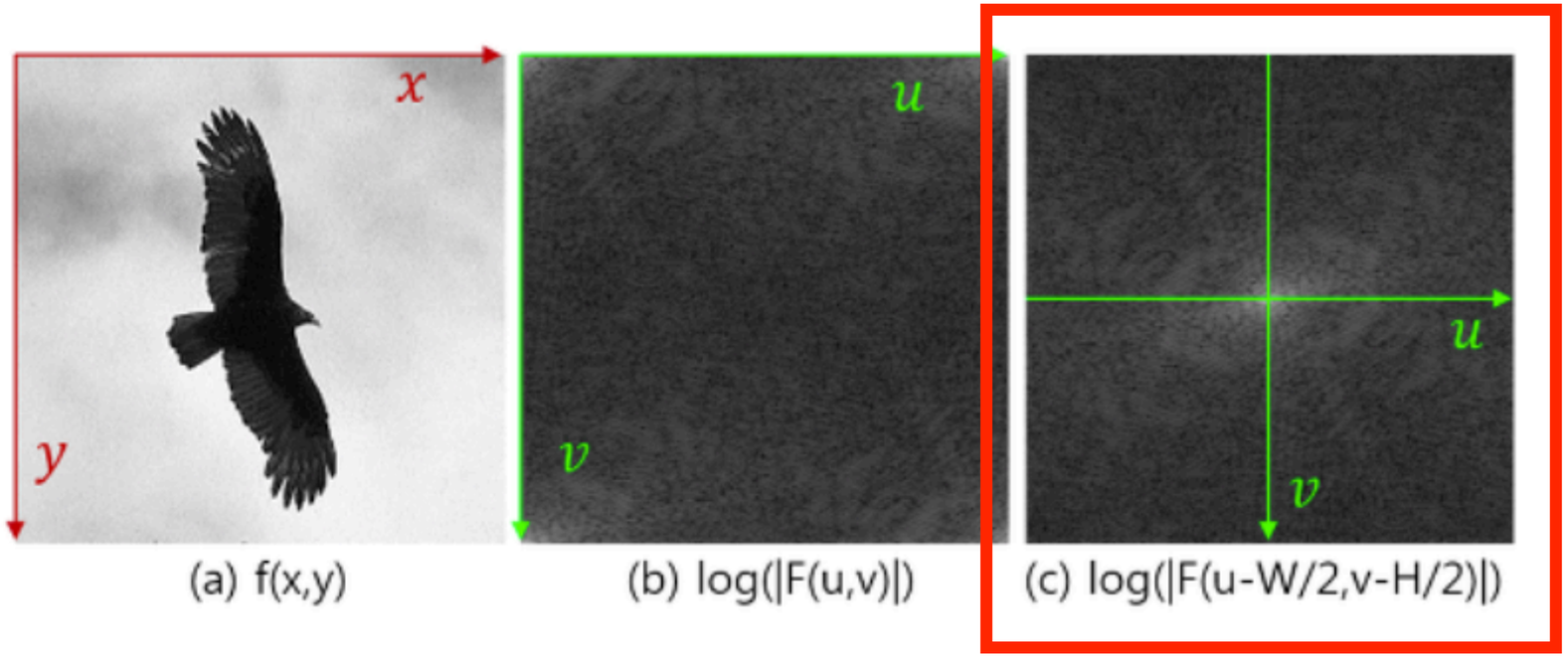
- 이산 푸리에 변환
  - 이미지는 이산(discrete) 신호
  - $$\mathcal{F}(I)(k,\ell) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} e^{-2\pi i\cdot\frac{jk}{M}} e^{-2\pi i\cdot\frac{j\ell}{N}} \cdot I(m,n),$$
    $$\text{for} \quad k = 0,\ldots,M-1, \quad \ell = 0,\ldots,N-1,$$
- $F(u,v)$는 x축 방향으로 주파수 u/W,
  y축 방향으로 주파수 v/H인 주기함수에 대한 계수

BACKGROUND

- 푸리에 변환
  - 스펙트럼과 페이즈(phase)를 얻음
  - $F(u,v) = R(u,v) + jI(u,v)$
    - $F(u,v)$의 크기 $|F(u,v)| = [(R(u,v))^2 + (I(u,v))^2]^{1/2}$ 를 스펙트럼이라 함
    - $F(u,v)$의 각도 $\phi(u,v) = tan^{-1}[\frac{I(u,v)}{R(u,v)}]$ 를 페이즈라 함

BACKGROUND

- 푸리에 스펙트럼
  - 해당 주파수 성분이 원 성분(이미지)에 얼마나 강하게 포함되는가



(a) f(x,y) (b) log(|F(u,v)|) (c) log(|F(u-W/2,v-H/2)|)

BACKGROUND

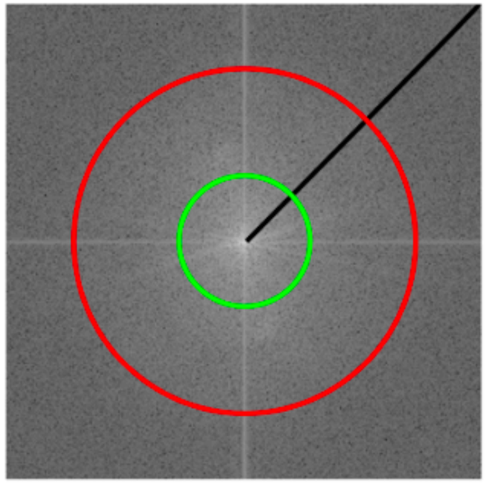
- AI: Azimuthal Integration (방위각 적분)
  - 2D 스펙트럼을 원점에서부터 반지름에 따라 합 또는 평균을 구한 것
    - 최대 반지름은 2D 스펙트럼 이미지의 한 변의 길이의 반
  - 결과는 1D plot

BACKGROUND

- Azimuthal Integration (방위각 적분)
  - 한 변의 길이가 M인 정사각 이미지를 가정
  - $$AI(\omega_k) = \int_0^{2\pi} \|\mathcal{F}(I)\left(\omega_k \cdot \cos(\phi), \omega_k \cdot \sin(\phi)\right)\|^2 \mathrm{d}\phi$$

    $$\text{for} \quad k = 0, \ldots, M/2 - 1\,,$$

BACKGROUND

- Azimuthal Integration (방위각 적분)
- 

- 이미지의 2D Power Spectrum을 1D Power Spectrum으로

BACKGROUND

- FID (Fréchet Inception Distance)
  - 진짜 데이터와 가짜 데이터의 특성 공간 상에서의 거리
  - Inception network(주로Inception V3)를 이용해
    진짜 데이터와 가짜 데이터의 특성을 추출
  - 두 집합에 대해 특성의 평균과 공분산을 구해 거리를 계산

ABSTRACT

ABSTRACT

- 생성적 합성곱 딥 뉴럴 네트워크
  - 합성곱(convolution) 기반 업샘플링 기법에 의존
  - 이미지나 비디오같은 논-스칼라 출력을 가짐
- 업샘플링 기법
  - 업-컨볼루션(up-convolution)
  - 전치된 컨볼루션(transposed convolution)

ABSTRACT

- 통상의 업샘플링 기법은 spectral 분포를 올바르게 생성하지 못함
  - 네트워크 구조와는 상관 없음
- 이를 이용하면 딥페이크(deepfakes)와 같은 생성된 데이터를 쉽게 검출
  - 벤치마크에서 100%의 정확도를 보임 (!)

ABSTRACT

- 이러한 한계를 극복하기 위해 저자들은 학습에 spectral 정형화(regularization, 정규화)를 사용
  - 분광-조화로운 GAN을 학습
  - 학습 안정성과 출력 품질을 향상시킴

INTRODUCTION

INTRODUCTION

- 가장 괄목할만한 생성적 신경망 구조는 다음 두 구조:
  - GAN(Generative Adversarial Networks)
  - VAE(Variational Auto Encoders)
- 두 방법은 학습 데이터 샘플의 분포의 잠재적 공간을 모사
  - 잠재적 공간이 주어지면 새로운 인공적 샘플을 만들 수 있음

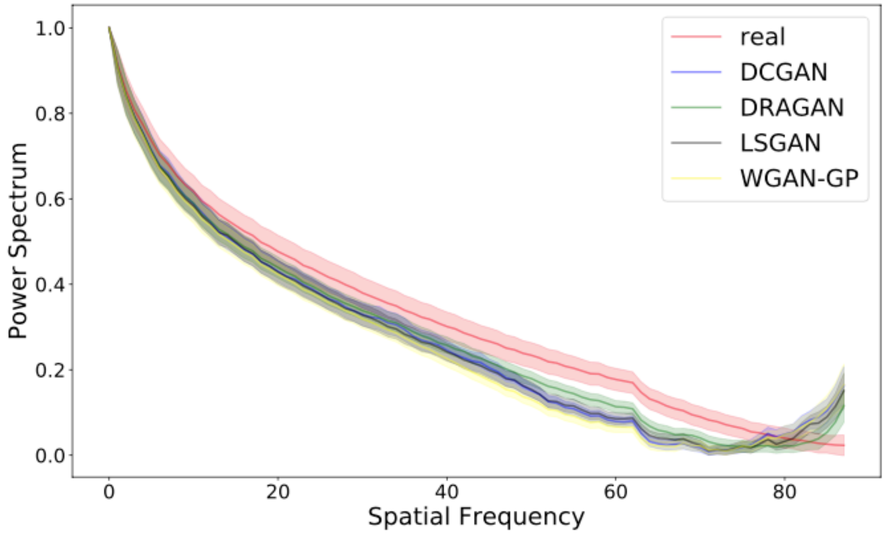
INTRODUCTION

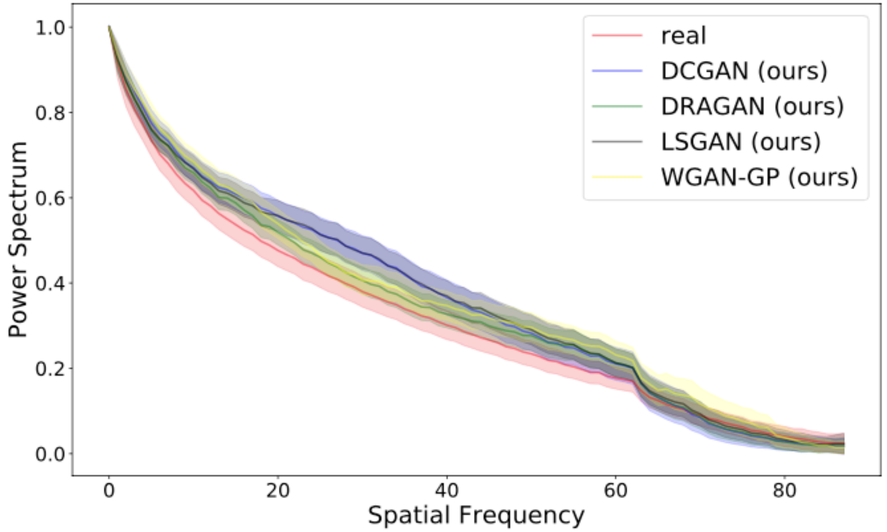
- 이러한 방법들은 기본 데이터 생성 원리를 따름:
  - 저차원 및 저해상도 잠재 공간을 고해상도 출력 공간으로 전환
    - 주로 1D —> 2D
  - 따라서 업스케일링 속성을 제공해야 함

INTRODUCTION

- GAN에서, 판별기 역시 주로 CNN을 사용
  - 임의의 확장으로부터 데이터(이미지) 분포를 따르는 보기에 좋은 출력을 만드는 것에는 충분
- 업스케일링 연산이 분광 속성을 (안 좋게) 변화시킴이 알려져 있음
  - 출력에 왜곡을 만들어냄
  - Jain, A. K. "Fundamentals of Digital Image Processing. 1989 Prentice HallEnglewood Cliffs." 88-102.

INTRODUCTION





- 통상의 업-컨볼루션 기법은 많은 공간 왜곡을 야기
- 저자들이 제안한 spectral 손실을 GAN 학습에 사용하면(아래) 기존의 GAN(위) 보다 실제 데이터에 가까운 spectral 분포를 보임

CONTRIBUTIONS

- 오늘날의 GAN이 하지 못한 spectral 분포를 근사할 수 있음
- 간단하지만 매우 정확한 감지기를 만들 수 있음
  - 공개 벤치마크 기준 DeepFake 감지 정확도 100%
- Spectral 왜곡을 해결할 수 있는 Spectral 정형화 방법을 제안함
- Spectral 정형화를 사용해 GAN 학습을 더욱 안정적으로 함
  - 출력의 품질도 향상

UP-CONVOLUTIONS

UP-CONVOLUTIONS IN GENERATIVE DONS

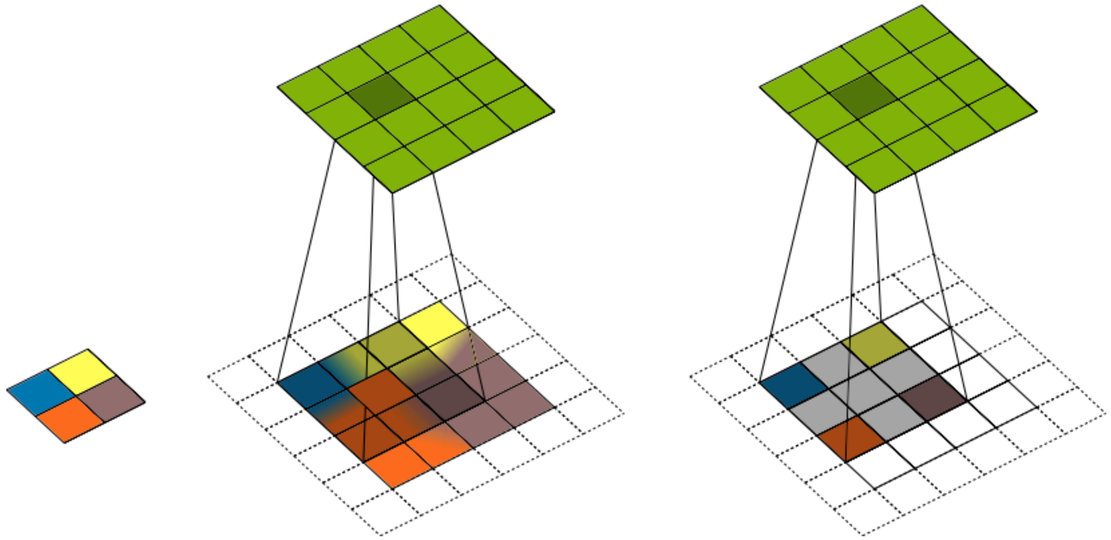
- GAN과 같은 생성적 신경망은 저차원에서부터 고차원 출력을 제공함
  - 가령, 노이즈에서 이미지
- 이러한 방법들은 어떠한 종류의 업-스케일링을 수행하게 됨

UP-CONVOLUTIONS IN GENERATIVE DONS

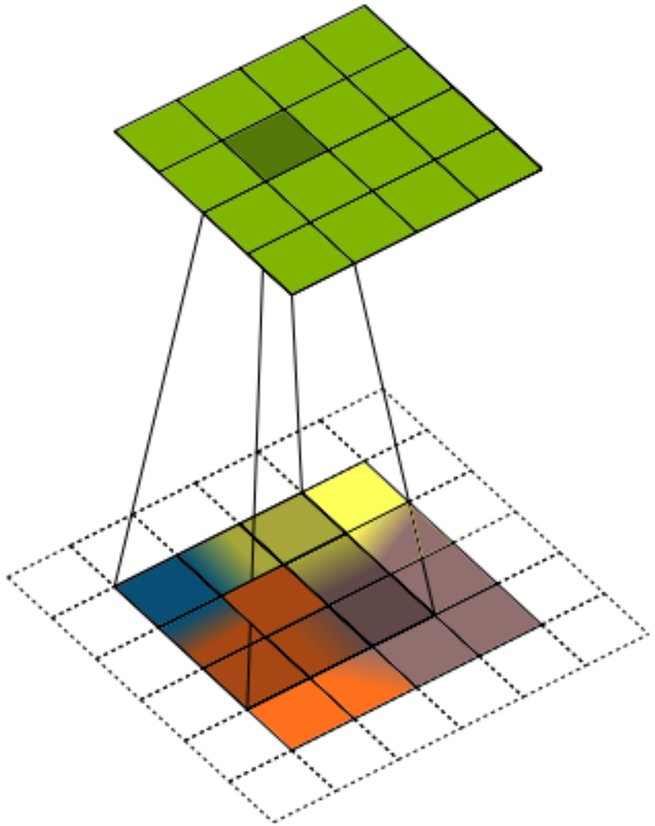
- 대표적인 것으로:
  - 보간법을 통한 업컨볼루션(up+conv)
  - 전치된 컨볼루션(transconv)

UP-CONVOLUTIONS IN GENERATIVE DONS

- (좌측) 저해상도 이미지: 2 x 2
- (가운데) 보간법을 통한 업컨볼루션(up+conv)
- (우측) 전치된 컨볼루션(transconv)

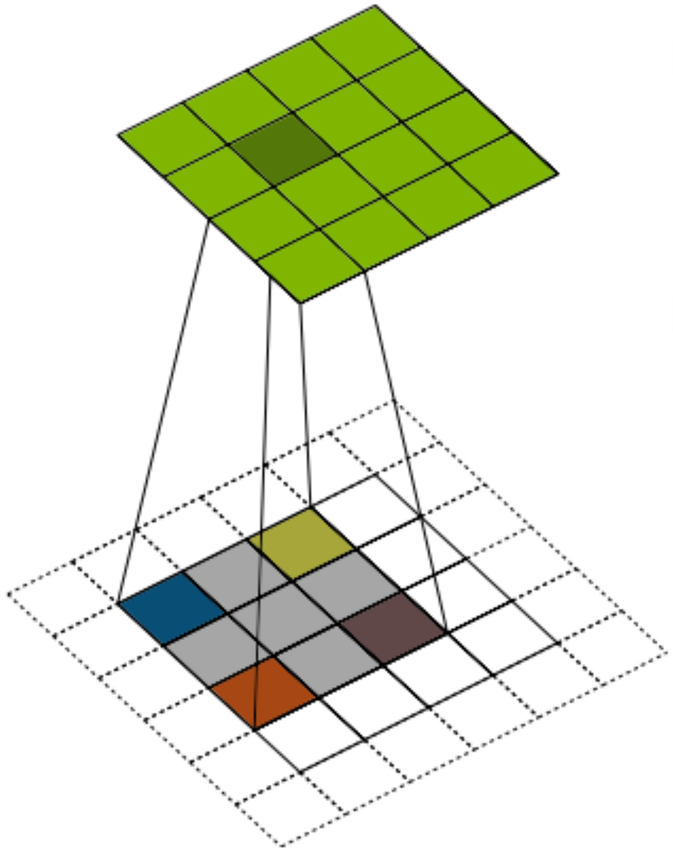


UP-CONVOLUTIONS IN GENERATIVE DONS



- 보간법을 통한 업컨볼루션(up+conv)
- 보간을 통해 채워넣음
  - Bi-linear
  - Nearest neighbor
- 그 후 컨볼루션
  - 2 x 2를 4 x 4로

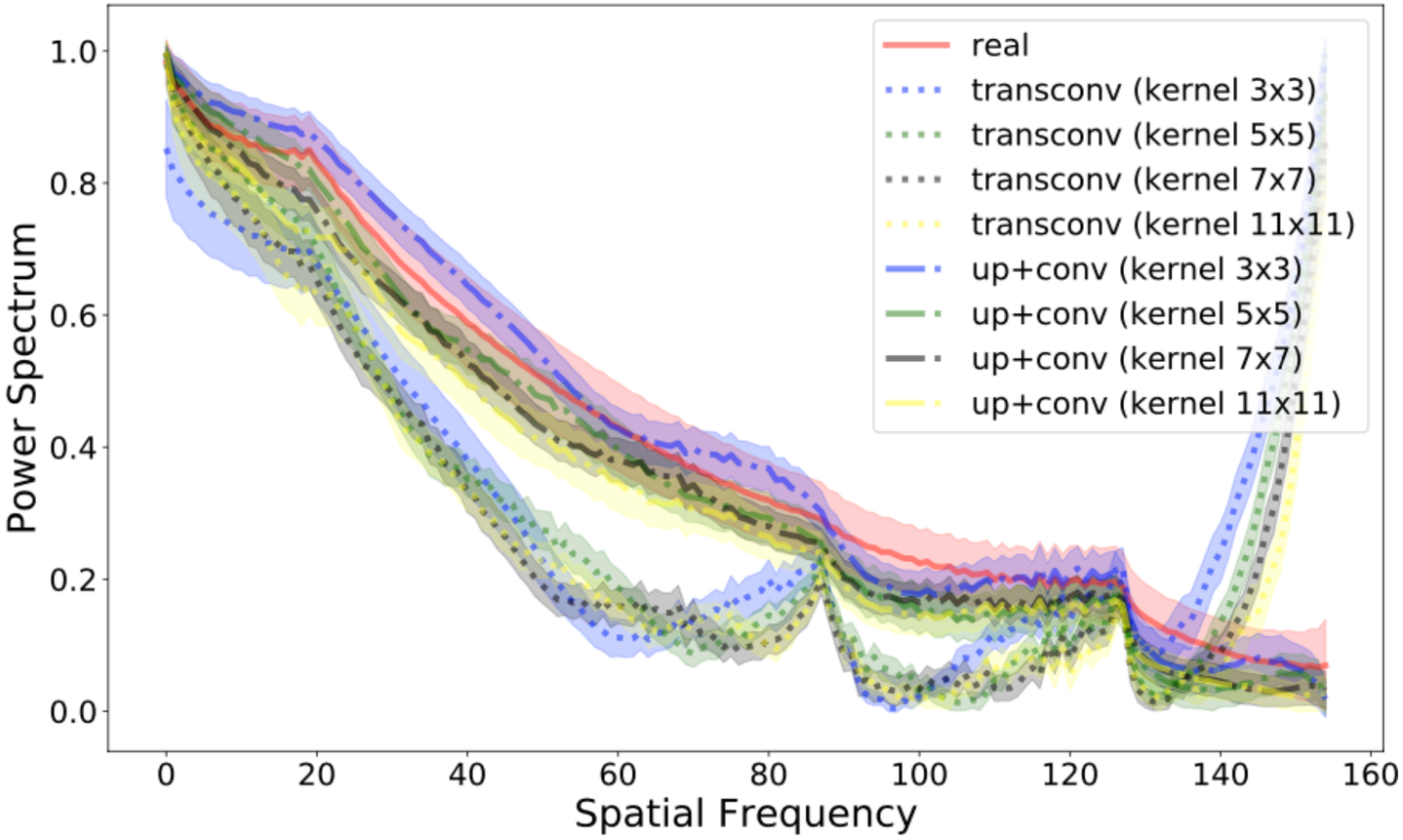
UP-CONVOLUTIONS IN GENERATIVE DONS



- 전치된 컨볼루션(transconv)
- 바늘방석(Bed of nails)으로 채워짐
  - 0으로 채워짐
- 그 후 컨볼루션
  - 2 x 2를 4 x 4로

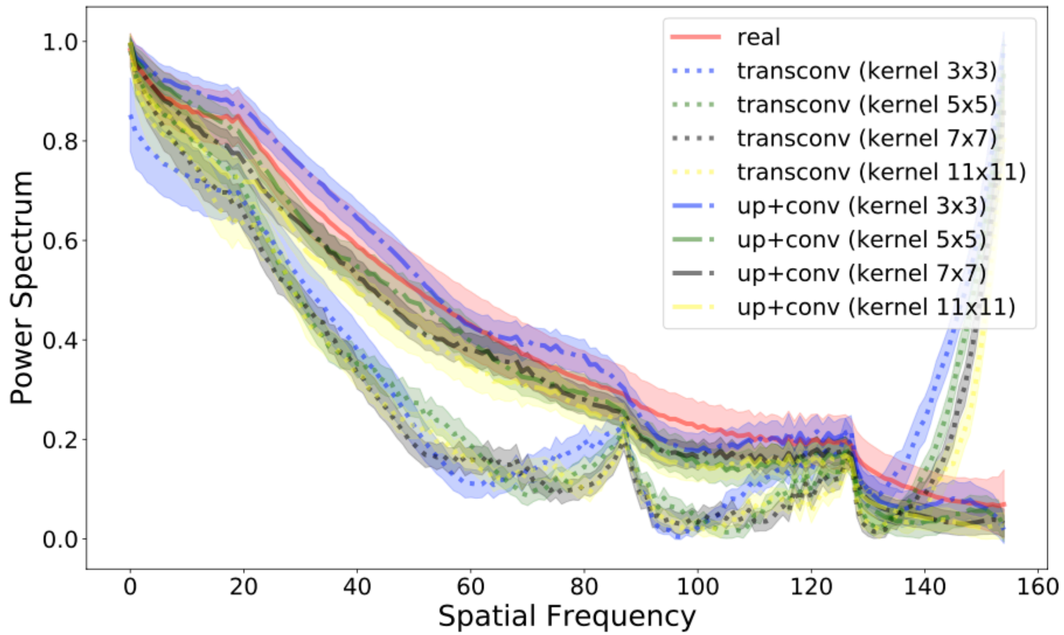
UP-CONVOLUTIONS IN GENERATIVE DONS

- 업-컨볼루션과 공간 주파수



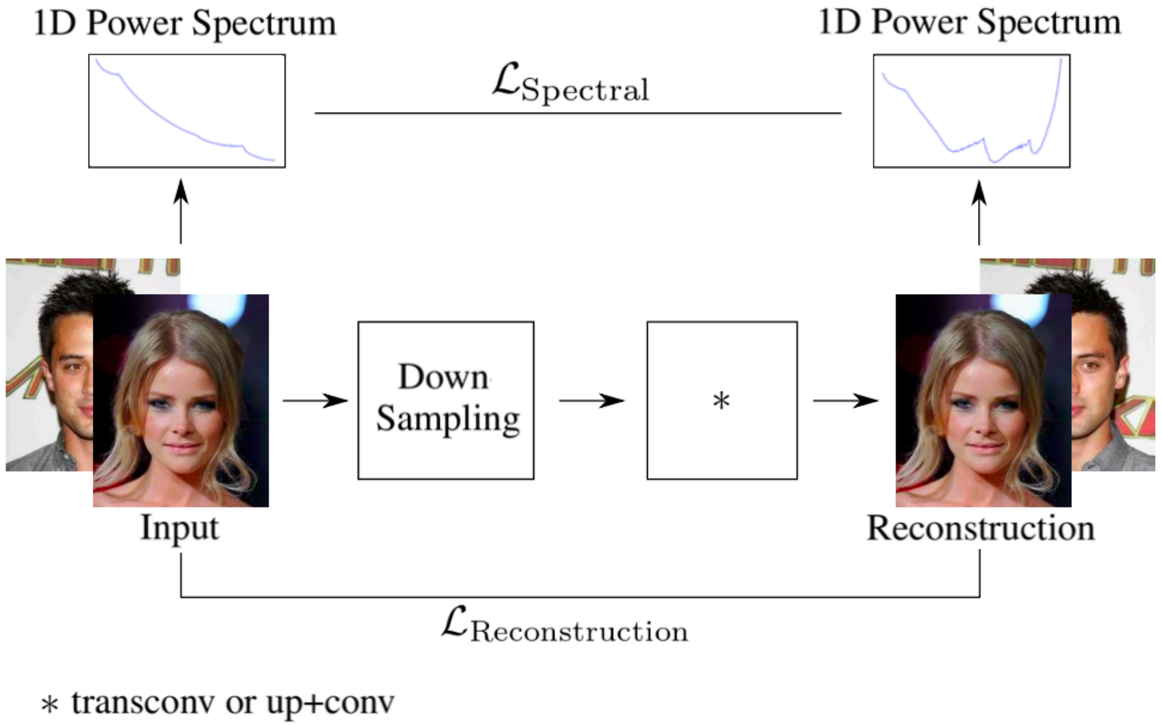
UP-CONVOLUTIONS IN GENERATIVE DONS

- 두 기법 모두 출력의 spectral 분포에 막대한 영향을 끼침
  - 전치된 컨볼루션은 막대한 고주파수 노이즈를 가짐
  - 반면, 보간법 기반 업컨볼루션은 고주파의 결여를 보임



UP-CONVOLUTIONS IN GENERATIVE DONS

- 이미지 출력의 Spectral 왜곡
  - 간단한 오토 인코더 (AE)를 사용해 2D 이미지에서 업-컨볼루션의 효과를 측정

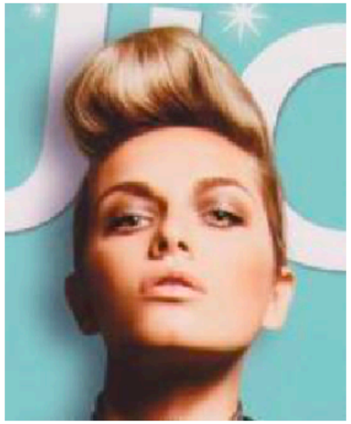
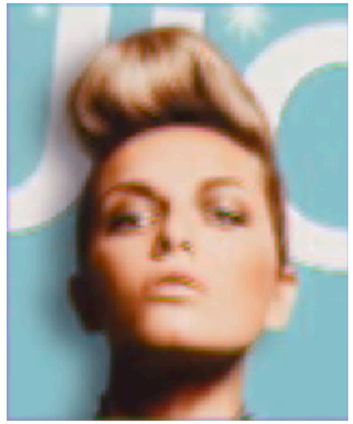
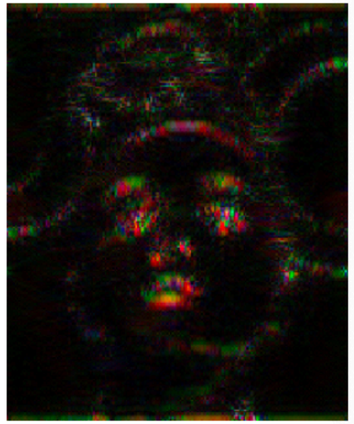


UP-CONVOLUTIONS IN GENERATIVE DONS

- (좌) 원본 이미지 / (중) AE 출력 / (우) 차이

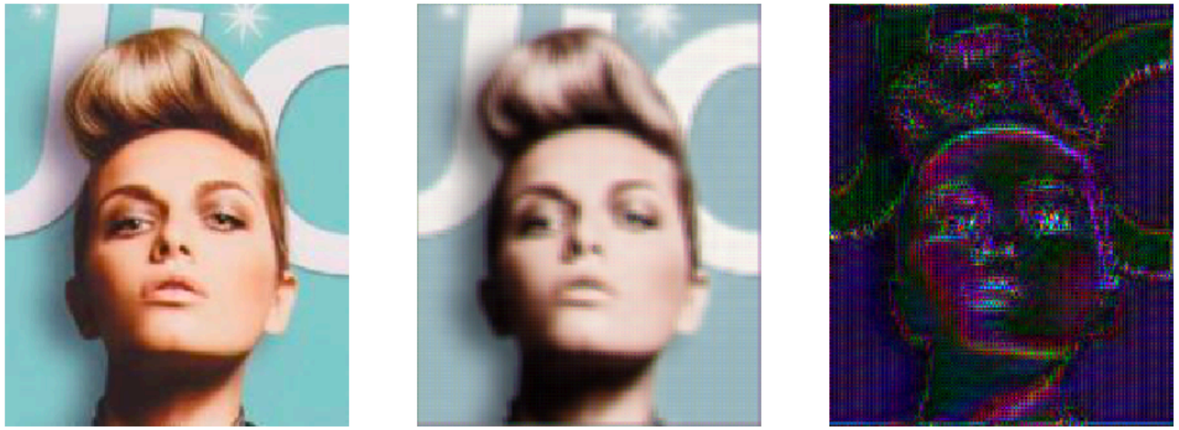
UP-CONVOLUTIONS IN GENERATIVE DONS

- Up+conv
  - 고주파수의 결여로 인한 흐려짐(blurring) 효과

UP-CONVOLUTIONS IN GENERATIVE DONS

- Transconv
  - 고주파수 노이즈



LEARN SPECTRAL

LEARNING TO GENERATE CORRECT SPECTRAL DISTRIBUTIONS

- 업-컨볼루션의 수행 시 이러한 spectral 왜곡을 보정할 수 있을까?
  - Spectral 정형화를 수행

LEARNING TO GENERATE CORRECT SPECTRAL DISTRIBUTIONS

- 통상의 생성적 네트워크 구조는
  - 이미지 공간에만 기반하므로
  - 올바른 spectral 왜곡을 담아낼 수 없음

SPECTRAL REGULARIZATION

- 저자들은 spectral 항(term)을 포함한 손실 함수를 제안
  - $L_{final} = L_{Generator} + \lambda L_{Spectral}$
  - λ는 spectral loss의 영향력을 결정짓는 하이퍼-파라미터

SPECTRAL REGULARIZATION

- $L_{final} = L_{Generator} + \lambda L_{Spectral}$
- $L_{Spectral}$ 은 binary cross entropy로 쉽게 할당할 수 있음
  - 생성된 출력 AI^{out} 과 진짜 이미지 샘플들의 평균 AI^{real} 사이

SPECTRAL REGULARIZATION

- $L_{Spectral}$ 은 binary cross entropy로 쉽게 할당할 수 있음
- $BCE(x) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\left(h(x_i;\theta)\right) + (1-y_i)\log\left(1-h(x_i;\theta)\right)$

SPECTRAL REGULARIZATION

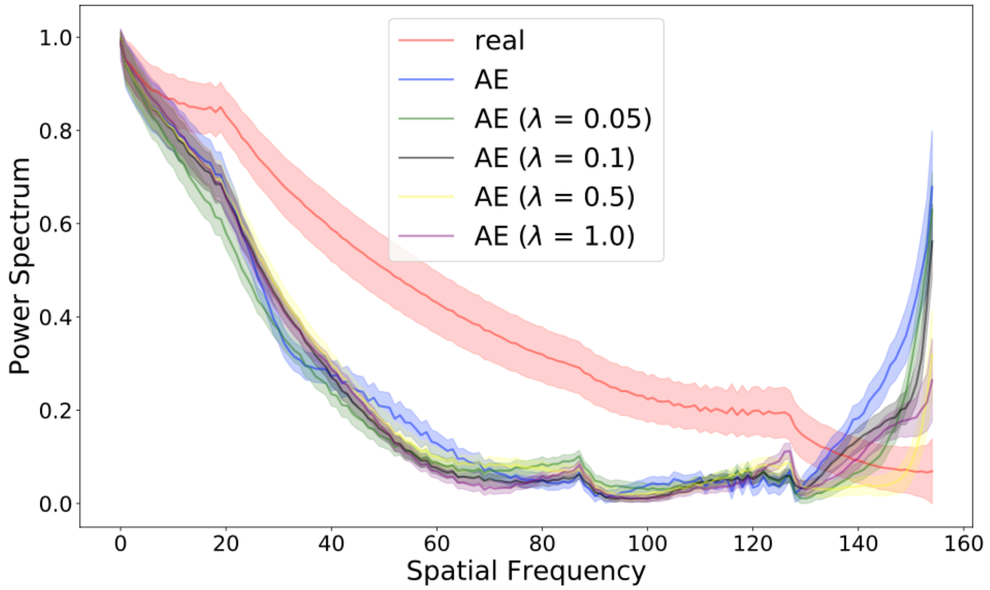
- 생성된 출력 AI^{out} 과 진짜 이미지 샘플들의 평균 AI^{real} 사이

$$\mathcal{L}_{\text{Spectral}} := -\frac{1}{(M/2-1)} \sum_{i=0}^{M/2-1} AI_i^{real} \cdot \log(AI_i^{out}) + (1 - AI_i^{real}) \cdot \log(1 - AI_i^{out})$$

  - M은 이미지 사이즈

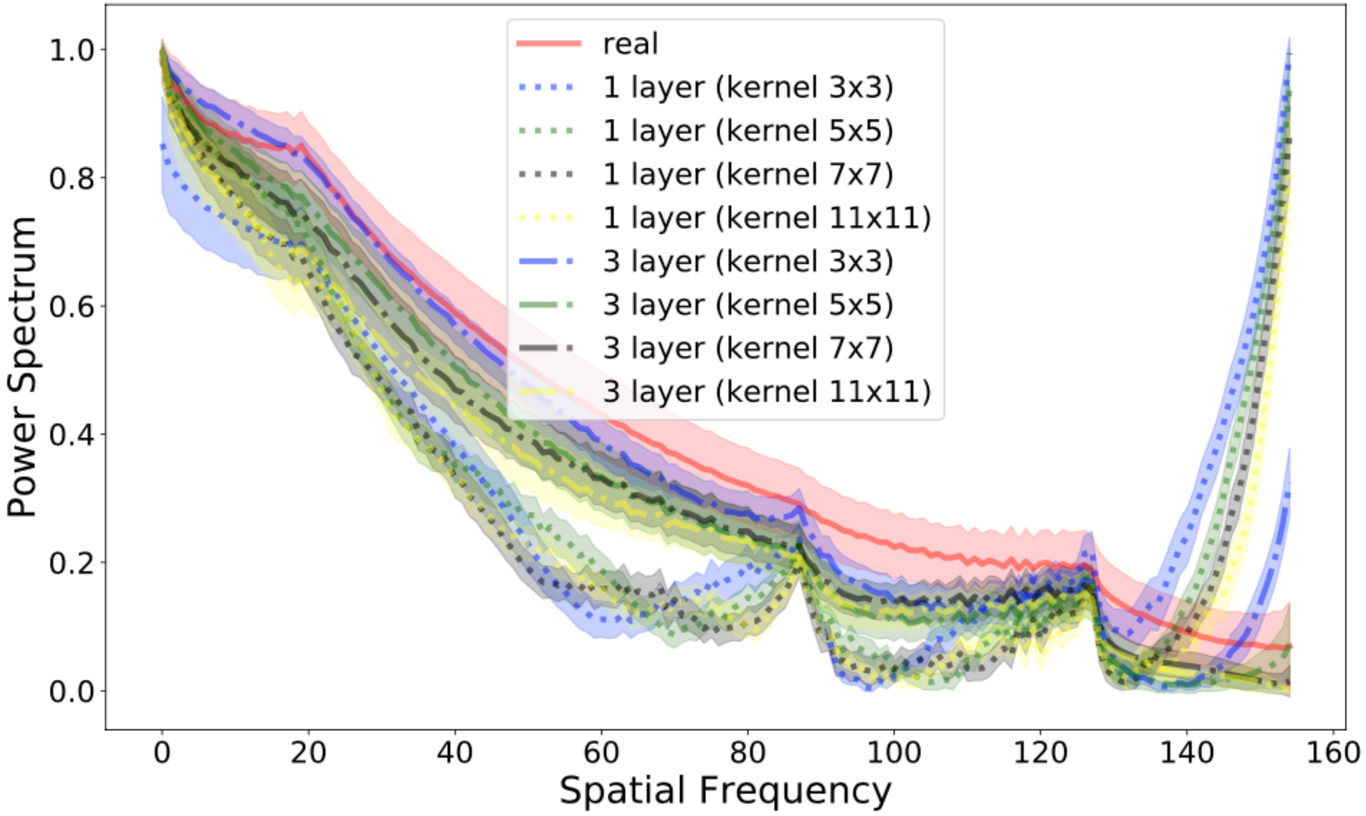
SPECTRAL REGULARIZATION

- λ에 따른 효과:
  - 가중치가 크더라도 별 효과 없음
  - 하나의 3 x 3 컨볼루션 레이어로는 효과 없음



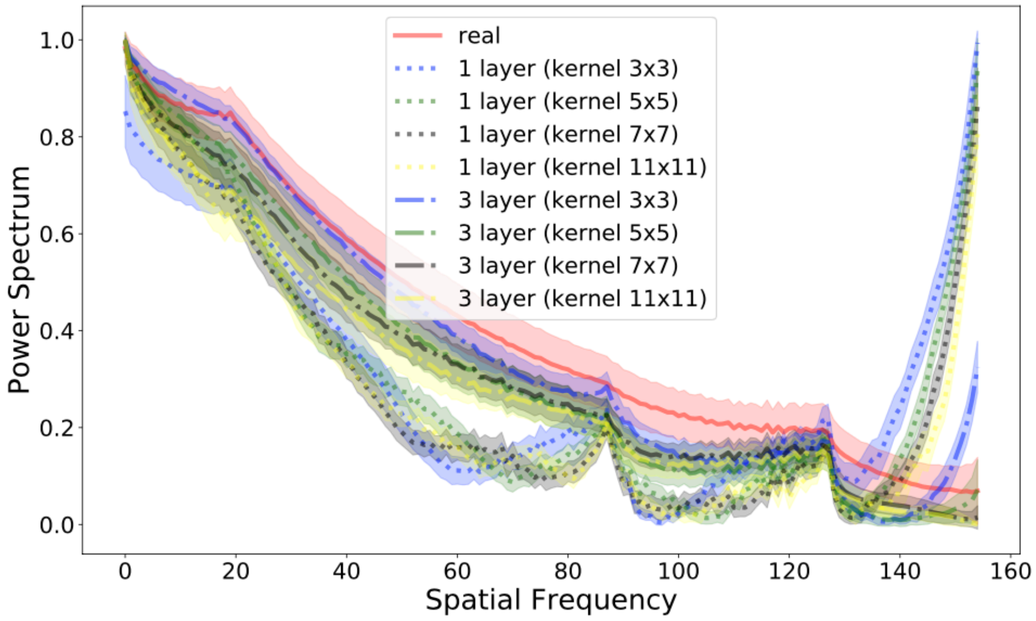
FILTER SIZES ON UP-CONVOLUTIONS

- Spectral loss + 필터 크기와 컨볼루션 레이어의 수 변화



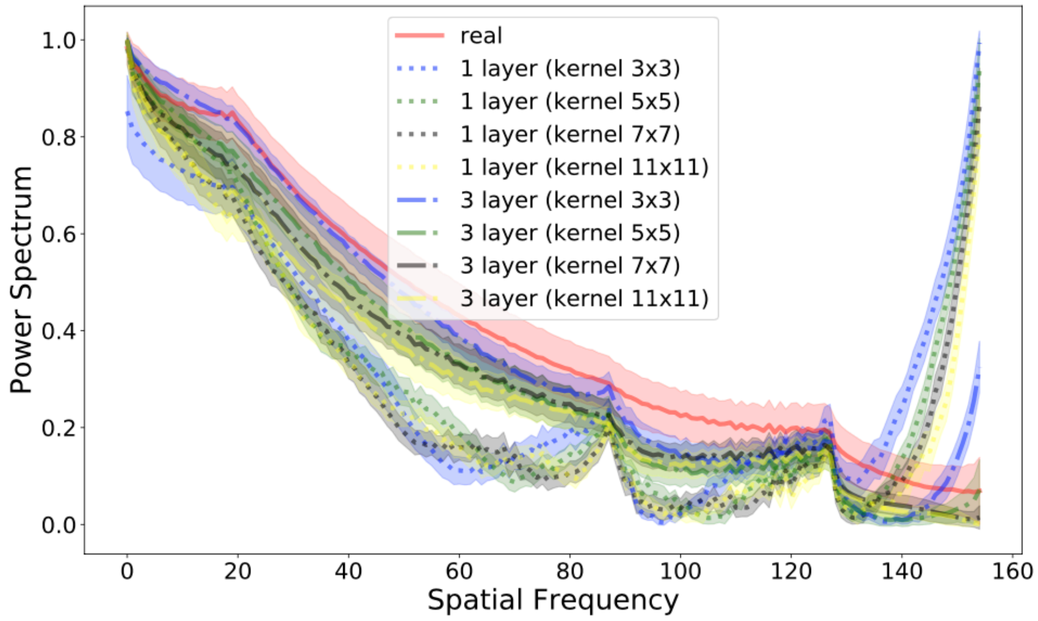
FILTER SIZES ON UP-CONVOLUTIONS

- 하나의 컨볼루션 레이어로는 spectral loss의 효과 없음
- 3 x 3 레이어에서도 효과 없음
- 더 많은(깊은), 더 큰 필터에서 효과 있음



FILTER SIZES ON UP-CONVOLUTIONS

- 필터 사이즈와 레이어 수의 결정이 중요
- 세 개의 5 x 5 컨볼루션 레이어에서 매우 우수한 spectral 왜곡 억제 효과



EVALUATION

EXPERIMENTAL EVALUATION

- 기존 유명한 GAN 구조로 생성한 가짜 이미지(얼굴)
  - 강한 spectral 왜곡이 있음을 보임
  - 가짜 혹은 인공 이미지로 검출할 수 있는 특성

EXPERIMENTAL EVALUATION

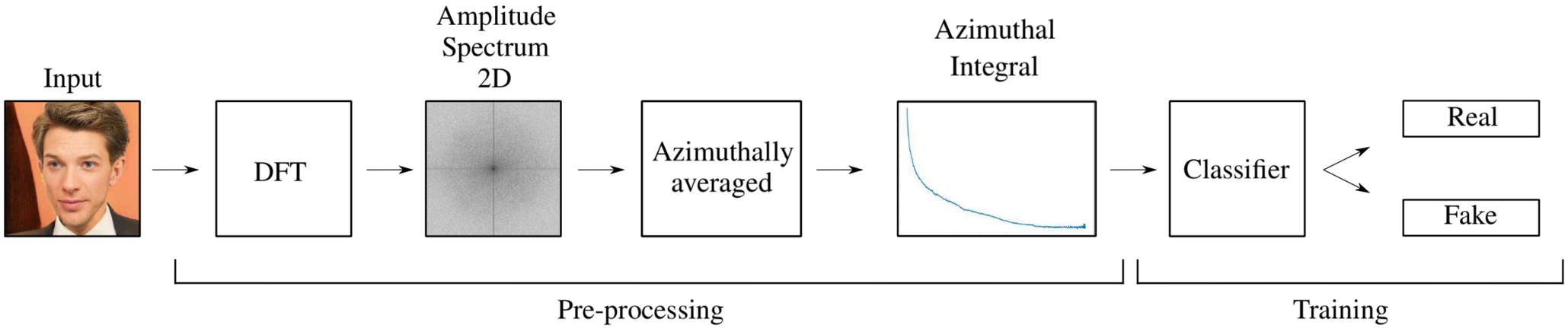
- 제안하는 spectral loss는 같은 데이터에 대해서
  - 주파수 영역에서 더 만족스러운 결과를 보임
- 또한, GAN 학습에서 더 안정적인 효과를 보임

DEEPFAKE DETECTION

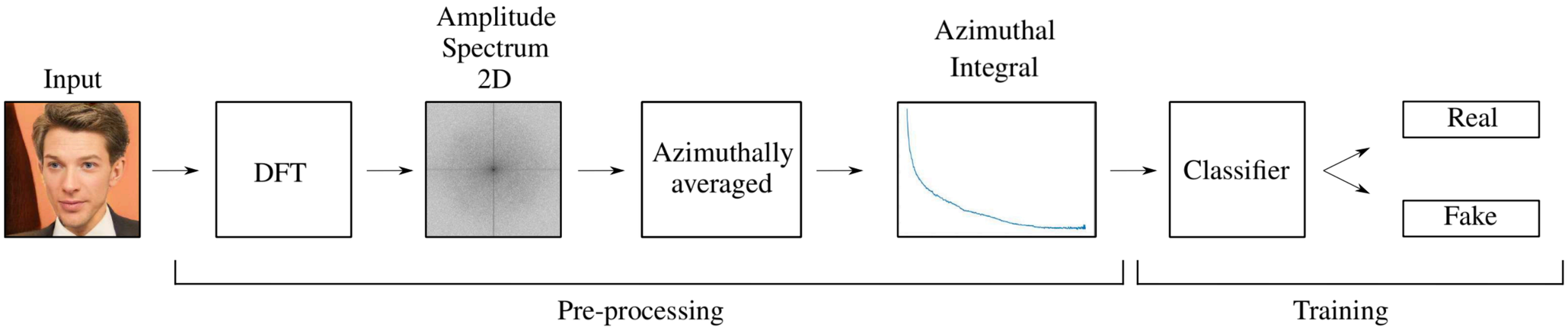
- 업-컨볼루션으로 인한 spectral 왜곡은 비록 SOTA인 GAN 구조에서도 매우 쉬운 가짜 이미지 검출이 가능하게 함
- Public 벤치마크에서 거의 완벽한 정확도로 생성된(가짜) 얼굴을 검출함

DEEPFAKE DETECTION

- 검출 방법의 도식
  - 이미지는 DFT(이산 푸리에 변환) 전에 흑백으로 변환



DEEPFAKE DETECTION



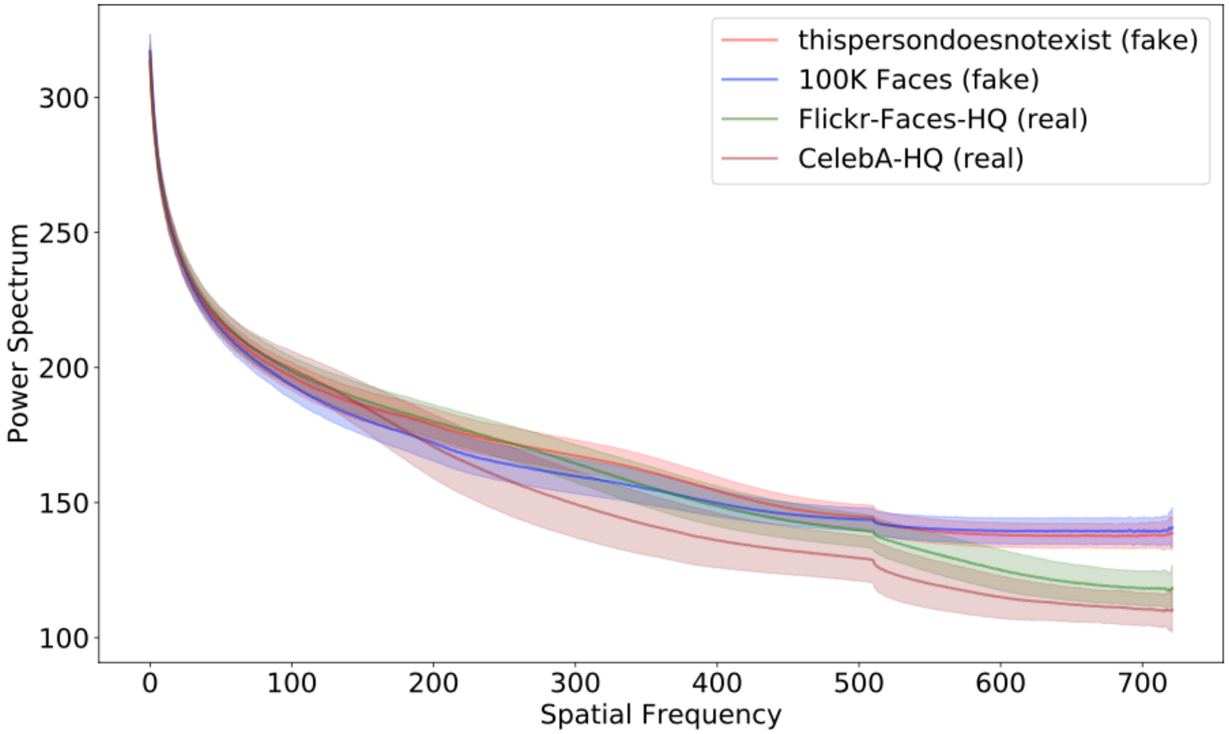
- 분류기
  - supervised에서는 SVM
  - unsupervised에서는 K-Means 사용

DEEPFAKE DETECTION

- 다른 이미지 크기를 조율
  - 1D 파워 스펙트럼을 정규화(normalization)
  - 1D 특성 벡터를 고정된 크기로 만듦

EXPERIMENTAL RESULTS

- Spectral 왜곡: 명백히 분류기가 real과 fake 구분 가능
  - 많은 데이터와 복잡한 구조에 기반한 DNN 검출기보다 월등히 성능이 좋음



EXPERIMENTAL RESULTS

- 테스트 정확도
  - A: 싱글 프레임에 대한 평가
  - B: 싱글 프레임 검출 —> 다수결 —> 전체 비디오에 대한 평가

	80% (train) - 20% (test)			
data set	method	# samples	supervised	unsupervised
Faces-HQ	ours	1000	100%	**82%**
Faces-HQ	ours	100	100%	81%
Faces-HQ	ours	20	**100%**	75%
CelebA	ours	2000	**100%**	**96%**
CelebA	[57]	100000	99.43%	-
CelebA	[39]	100000	86.61%	-
FaceForensics++	ours^A	2000	85%	-
FaceForensics++	ours^B	2000	**90%**	-

EXPERIMENTAL RESULTS

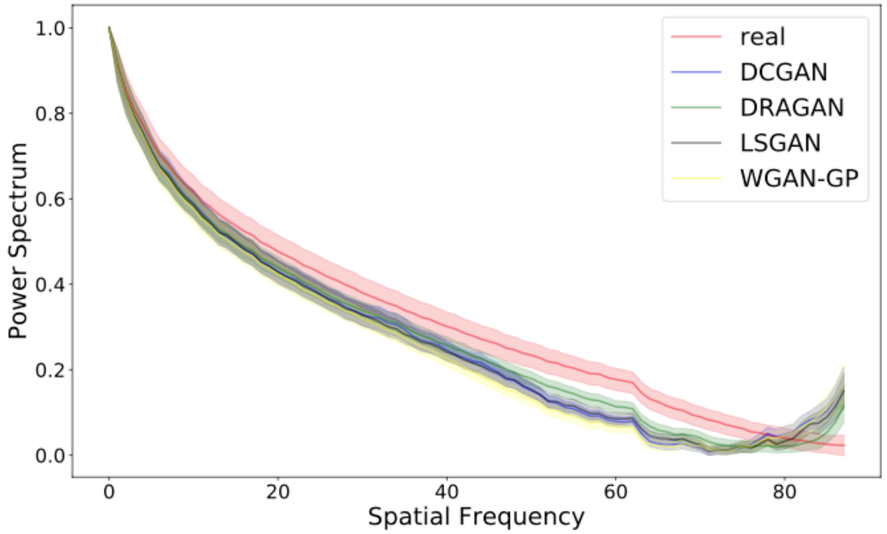
- [57] N. Yu, L. Davis, and M. Fritz. Attributing fake images to gans: Learning and analyzing gan fingerprints. In Inter- national Conference on Computer Vision (ICCV), October 2019.
- [39] F. Marra, D. Gragnaniello, L. Verdoliva, and G. Poggi. Do gans leave artificial fingerprints? CoRR, abs/1812.11842, 2018.

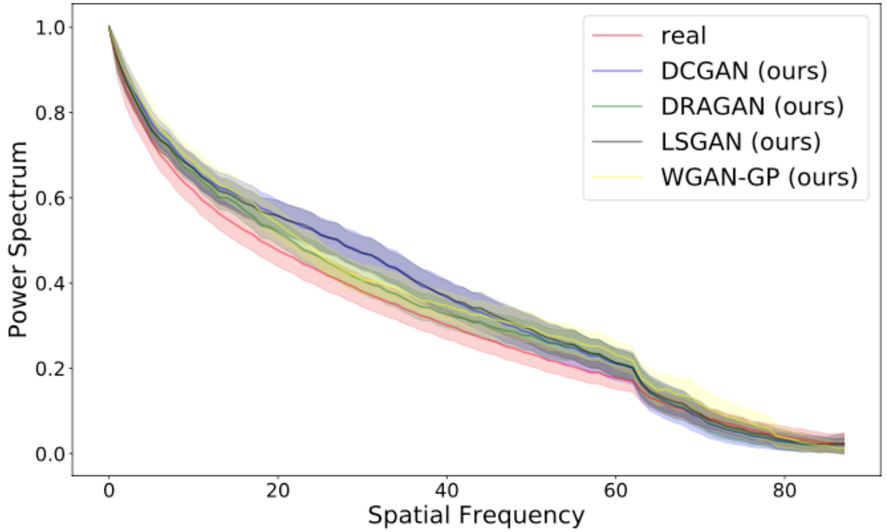
	80% (train) - 20% (test)			
data set	method	# samples	supervised	unsupervised
Faces-HQ	ours	1000	100%	**82%**
Faces-HQ	ours	100	100%	81%
Faces-HQ	ours	20	**100%**	75%
CelebA	ours	2000	**100%**	**96%**
CelebA	[57]	100000	99.43%	-
CelebA	[39]	100000	86.61%	-
FaceForensics++	ours^A	2000	85%	-
FaceForensics++	ours^B	2000	**90%**	-

APPLYING SPECTRAL REGULARIZATION

- 정형화를 포함해 다시 실험을 진행
  - 생성기 손실에 Spectral loss 항을 추가
  - 1000개의 실제 샘플에서 AI^{real} 을 구함
  - 마지막 업-컨볼루션에서 컨볼루션 레이어를 5 x 5 커널을 가진 세 필터 레이어로 변경

APPLYING SPECTRAL REGULARIZATION

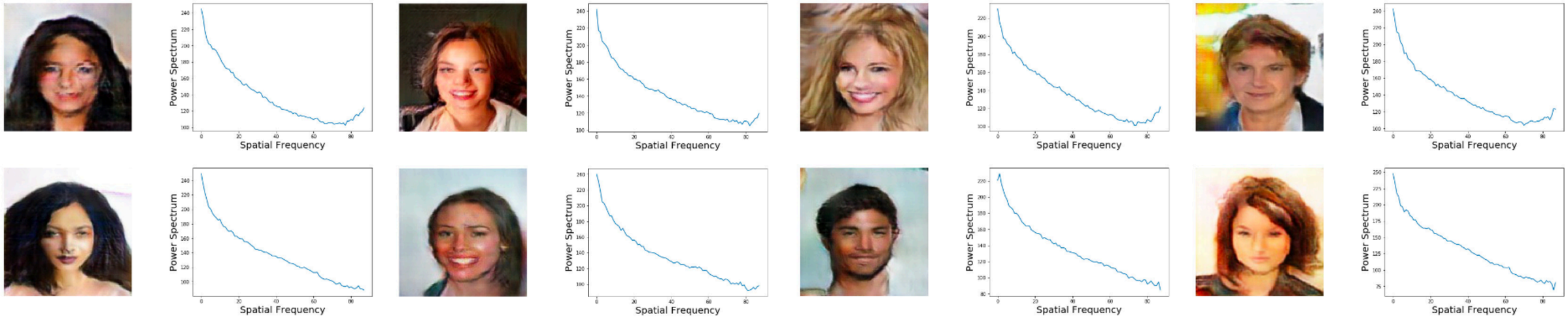




- 저자들이 제안한 spectral 손실을 GAN 학습에 사용하면(아래) 기존의 GAN(위) 보다 실제 데이터에 가까운 spectral 분포를 보임

APPLYING SPECTRAL REGULARIZATION

- 실제 생성 결과물도 질적으로 좋은 결과를 보임
  - 위: 기본 GAN / 아래: spectral 정형화 기법을 적용



(a) DCGAN. (b) DRAGAN. (c) LSGAN. (d) WGAN.

POSITIVE EFFECTS OF SPECTRAL REGULARIZATION

- 보기에 진짜같을 뿐만 아니라, 주파수 영역도 흉내낼 수 있음
  - 진짜 분포에 한 발짝 가까이 다가설 수 있음

POSITIVE EFFECTS OF SPECTRAL REGULARIZATION

- 흥미로운 사이드-이펙트
  - 실험에 따르면 spectral loss 항을 가진 GAN이 모드 붕괴(mode-collapse)에 빠질 위험이 적음
    - 모드 붕괴: 진짜 데이터의 분포의 매우 좁은 영역에 갇힘
  - 안정적으로 학습
  - 더 잘 수렴함

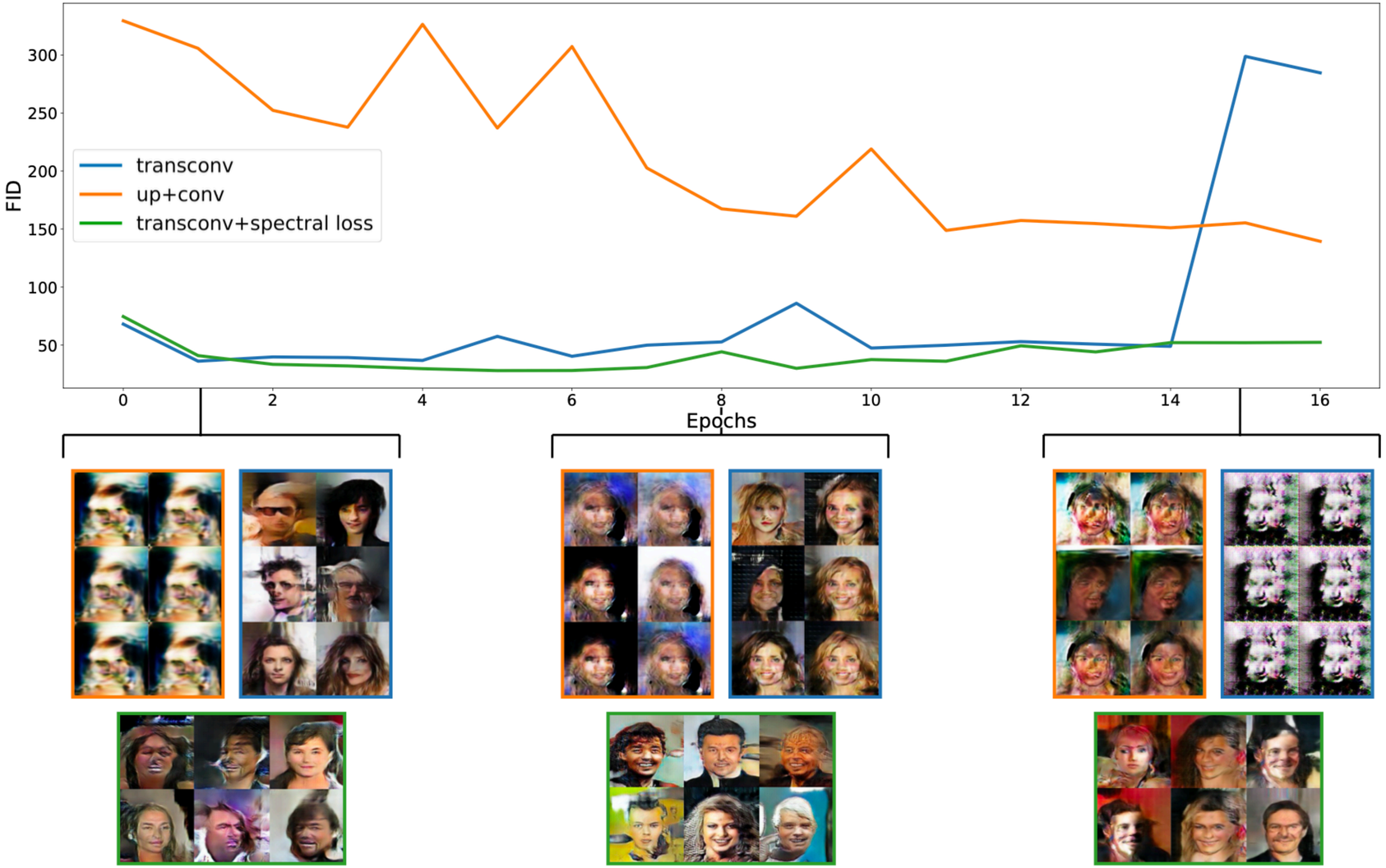
POSITIVE EFFECTS OF SPECTRAL REGULARIZATION

- FID와 GAN 출력 사이의 상관관계
  - DCGAN, CelebA 데이터셋
  - FID로 평가

POSITIVE EFFECTS OF SPECTRAL REGULARIZATION

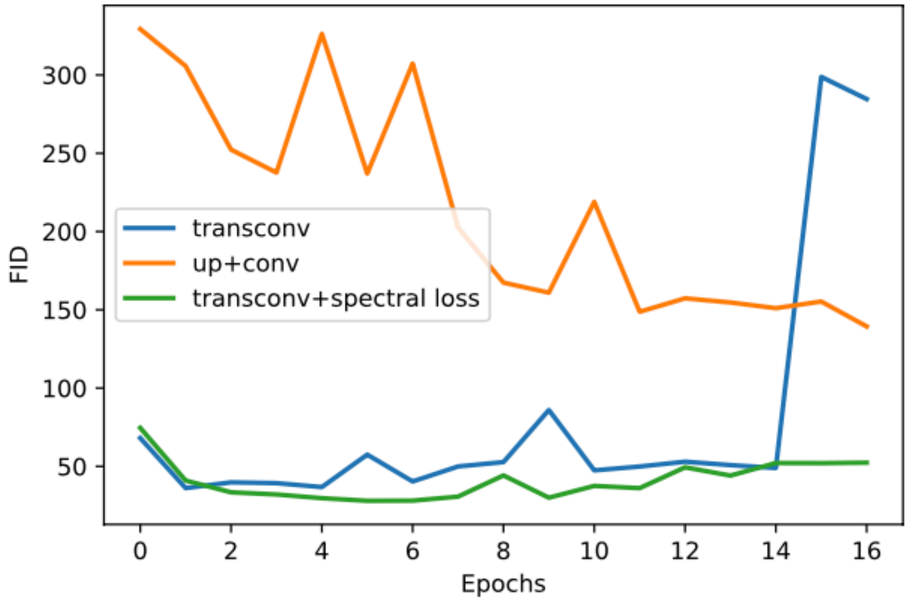
- FID와 GAN 출력 사이의 상관관계
  - 낮은 FID 점수는 다양성과 좋은 결과와 연관
  - 높은 FID 점수는 낮은 품질의 출력과 모드 붕괴와 연관

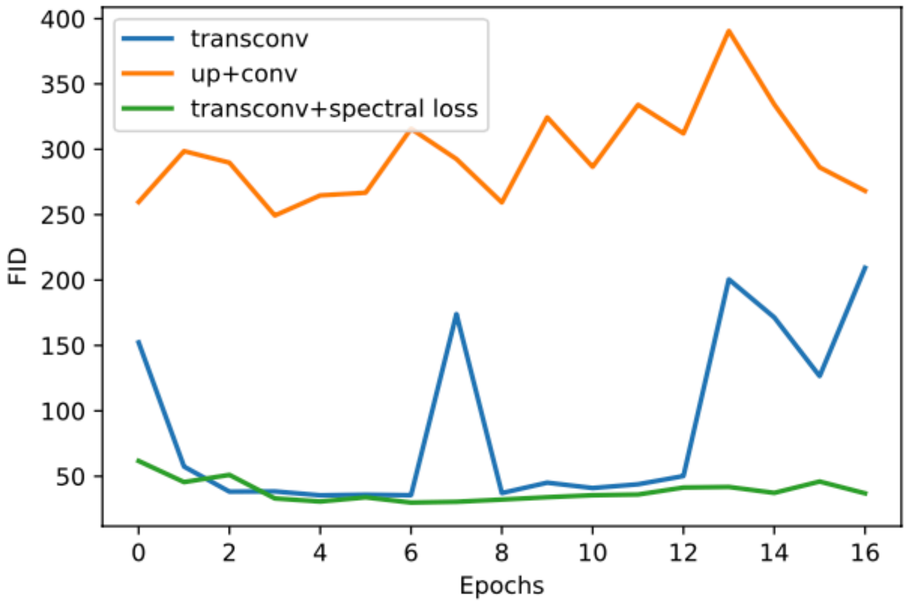
POSITIVE EFFECTS OF SPECTRAL REGULARIZATION



POSITIVE EFFECTS OF SPECTRAL REGULARIZATION

- Up+conv 버전의 DCGAN, LS-GAN은 실험 기간 동안 향상에 실패
- Transconv 버전은 수렴하지만 불안정
- Transconv+spectral loss는 낮고 안정적인 FID 점수를 획득 (좌: $\lambda = 2$ / 우: $\lambda = 0.5$ 모두)





CONCLUSION

DISCUSSION AND CONCLUSION

- 유명한 SOTA (state of the art) 컨볼루션 생성적 네트워크, GAN 생성기들이 spectral 왜곡을 보임
- 생성된 가짜 샘플 검출이 쉬움
- 역이용하여, 데이터 생성 향상에 도움이 됨
  - Spectral loss 항을 추가

DISCUSSION AND CONCLUSION

- Spectral loss 항
  - 생성기의 마지막 컨볼루션은 적어도 5 x 5 크기는 되어야 효과를 봄
  - 실험적으로, GAN의 안정성에도 도움이 됨을 보임
    - 이론적 분석이 추가 필요
    - 아마 고주파수 노이즈에 대해 인식하게 함으로써 판별기와 생성기에 영향을 주지 않았나 추정

WATCH YOUR UP-CONV

CNN BASED GENERATORS ARE DOOMED