

(H. Kim) Deep image prior

🕒 작성일 시	@November 16, 2021 10:12 AM
☰ Keywords	Deep Image Prior Image Restoration
👤 Reviewer	김하연
📄 Paper	https://sites.skoltech.ru/app/data/uploads/sites/25/2018/04/deep_image_prior.pdf
📅 발표일 자	@September 10, 2021
📎 발표영 상	
🕒 최종 편 집	@January 12, 2022 7:43 AM
☑ Prof.	☑
📎 PPT	
☰ Q&A Link	
☰ 비고 (추가 노 트)	
📎 추가자 료	
☑ Reviewed	☑
☰ 열	
☰ 속성	
☰ 속성 1	

211114 ~ 211115 DIP 내용정리(아이패드)
211116 초안 작성
211117 큰 틀 잡기, 세부 내용 작성

0. Abstract

In this paper, we show that the structure of a generator network is sufficient to capture a great deal of low-level image statistics prior to any learning.

→ 이전 모델들의 경우, 많은 example images로 훌륭한 performance를 보임. 하지만, 이 논문에서는 generate network의 구조 자체가 low-level image statistics를 capture하는데 충분하다고 보임.

→ 이는 훈련 data set이 많지 않아도 된다는 장점을 가지지만, 상대적으로 오랜 시간이 걸림.

- randomly-initialized neural network = handcrafted prior in standard inverse problem(네트워크 자체가 prior로 사용 가능)

Apart from its diverse applications, our approach highlights the inductive bias captured by standard generator network architectures.

Inductive bias(귀납적 편향) : 학습 시에는 만나보지 않았던 상황에 대하여 정확한 예측을 하기 위해 사용하는 추가적인 가정 (additional assumptions)을 의미. 즉, 일반적인 학습 데이터를 넘어 예기치 못한 데이터가 들어왔을 때에도 일반화할 수 있는 능력을 길러줌.

1. Introduction

- ConvNets의 경우 대표적으로 inverse image reconstruction problem(like denoising, single-image super-resolution)을 해결하고 있다. ConvNets은 큰 dataset의 이미지들을 필요로 하고 있다. 그들의 능력은 실제 이미지 priors를 학습하는 능력에 있어서 기인한다고 생각할 수 있으나, 이러한 단순히 '학습'만으로는 deep network의 훌륭한 수행 능력을 설명하기엔 불충분!
- 즉, 일반화는 네트워크의 구조가 데이터의 구조와 **"resonate"** 하도록 요구한다.

▼ 논문에서 왜 "resonate"를 강조했을까?

고민중.. 😞

- 그래서, 이 논문에서는 훈련과 별개로 convolutional image generator의 구조 만으로 많은 image statistics를 포착할 수 있다는 것을 보여줄 예정 → image restoration problem문제 해결 (generator network를 하나의 degraded image와 match시켜서 네트워크 자체가 restored image의 parameterization으로서 역할을 하는 것을 보일 예정)

▼ 그래서 가장 큰 이 논문의 필요성은?

네트워크는 데이터로부터 그 어떠한 학습도 받지 않음. 대신, 네트워크의 가중치는 항상 randomly 초기화된다. ⇒ **유일한 prior information = 네트워크 구조 itself!**

2. Method

- generator/decoder networks $x = f_{\theta}(z)$ where code vector z , image x
 - random distribution으로 부터 실제 이미지를 샘플링 하는데 사용 가능(corrupted observation x_0 에 기인하는 distribution에 focus할 예정!)
 - 이 함수의 목표 : 복잡한 확률분포인 $p(x)$ 를 바로 구하기 힘들므로, 간단한 확률분포인 $p(z)$ 를 이용할 수 있다. 이때, $f_{\theta}(z)$ 가 x 로 부터 기인한 확률분포인 $p(x)$ 와 가까워 질수록 선명한 사진을 얻을 수 있다
- DIP에서는 신경망 네트워크를 parametrization으로 해석($x = f_{\theta}(z)$, image $x \in \mathbb{R}^{3 \times H \times W}$) → 특히, DIP에서는 U-Net type "hourglass" architecture with skip-connections 사용 예정(x, z 크기 동일)
- 이러한 방식을 energy minimization problem으로 접근

$$x^* = \min E(x; x_0) + R(x)$$

where $E(x; x_0)$ is a task-dependent data term, x_0 the noisy/low-resolution-occluded image, and $R(x)$ a regularizer.

우리는 iteration 을 통해 에너지가 점점 낮아 지는 곳으로 자리를 옮겨가며, 에너지 함수의 일차 미분 값이 0(기울기 0, 함수의 최솟값) 이 되는 극점을 찾는다. ≡ gradient descent

- 여기서 $R(x)$ 를 다음과 같이 대체한다 :

$$\theta^* = \operatorname{armin} E(f_{\theta^*}(z); x_0), x^* = f_{\theta^*}(z)$$

여기서 minimizer θ^* 는 random initialization of the parameters으로부터 시작하는 경사 하강법 같은 optimizer를 이용하여 획득 가능

*** (이해가 안 간 부분 $\pi\pi$) $R(x) = 0$ for all images that can be produced from z by a deep ConvNet of a certain architecture, and $R(x) = +\infty$ for all other signals**

- Optimization

(1)

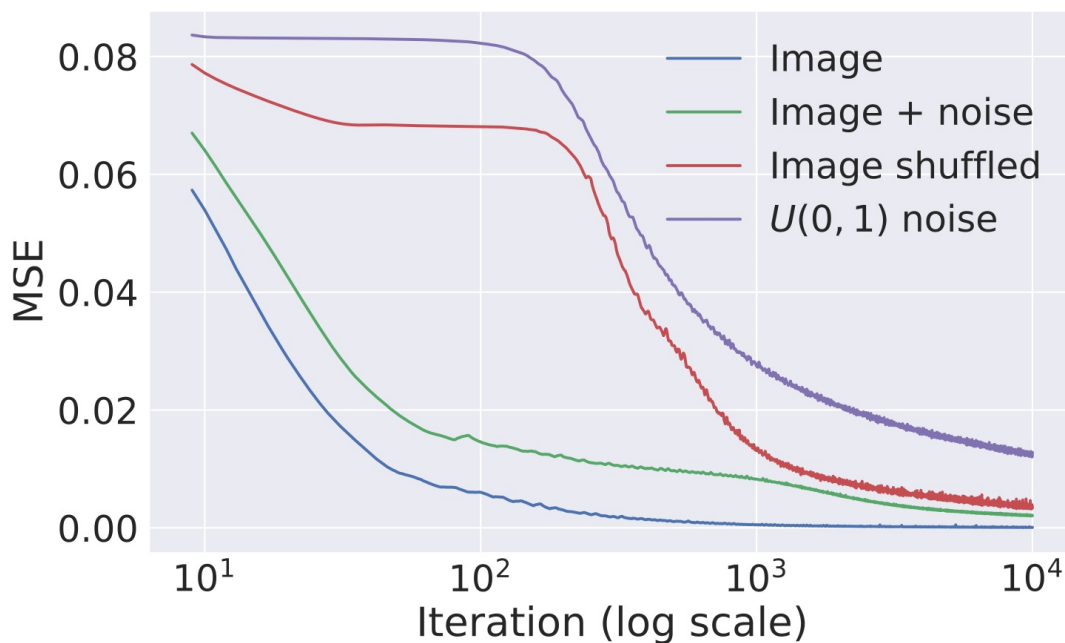
$$E(x; x_0) = \|x - x_0\|^2$$

(2)

$$\min ||f_{\theta}(z) - x_0||^2$$

즉, 우리는 만들어낸 이미지(output)인 $f_{\theta}(z)$ 를 최대한 x_0 에 근사시키는 것을 목표로 한다.(차이 최소화, 최대한 원본에 가까운 깨끗한 이미지) → 에너지가 최소화 되는 방향으로 움직인다.

The parameterization offers high impedance to noise → 노이즈에 강한 저항을 가지므로 훈련 필요없이 네트워크 자체가 restored image의 parameter로서 노이즈를 제거하는데 큰 역할을 하는 것을 알 수 있다.



▼ Gradient Descent(경사하강법)

- 함수 값이 낮아지는 방향으로 독립 변수 값을 변형시키며 최종적으로 최소 함수 값을 갖도록 하는 독립 변수 값을 찾는 방법(즉, 함수의 최솟값을 찾는 문제)
- gradient descent는 함수의 기울기(gradient)를 이용해 x의 값을 어디로 옮겼을 때 함수가 최솟값을 찾는지 알아보는 방법 → 기울기가 양수(증가)일때는 x를 음의 방향으로 옮기고 기울기가 음수(감소)일때는 x를 양의 방향으로 옮겨야 최솟값을 찾을 수 있다.

$$x_{i+1} = x_i - \alpha \frac{df}{dx}(x_i)$$

- DIP에서는 이 방식으로 minimizer θ^* 을 구하고 있다.

논문을 마치며 : 딥 러닝에서는 hand-craft priors보다 학습을 하는 것이 이미지 복구에 큰 역할을 한다는 기존의 학설과 반대로, 네트워크 구조 그 자체만으로도 훌륭한 이미지 복구 역할을 한다는 것을 보여주었다.(비록 느리지만,,) 이는 앞으로의 새로운 딥러닝 architecture의 중요성을 강조하며 마무리한다.

▼ 나를 위한 용어 정리

1. HOG descriptor : Histogram of Oriented Gradients 의 줄임말로 image의 지역적 gradient를 해당 영상의 특징으로 사용하는 방법. HOG는 이미지 경계의 기울기 **벡터 크기 (magnitude)**와 **방향(direction)**을 **히스토그램으로 나타내** 계산.
2. TV norm : total variation norm
3. MSE : Mean Squared Error