cartoonGAN 논문 요약

Abstract

실제 현실의 사진을 cartoon style images 로 변환하는 것을 목표로 한다.

이 논문에서 나오는 solution은 최근 stylize images에 널리 사용되는 학습방법에 기초한다. (GAN)

하지만 현존하는 방법은 cartoonization에 만족스러운 결과를 도출해내지 못한다.

그 이유는

(1) cartoon style 은 높은 level의 간단함과 추상적인 특징을 가지며

(2) cartoon image는 clear edges 와 부드러운 color shading 과 간단한 texture을 가지는데 이는 현존하는 texture-descriptor-based loss functions 으로는 해결하기 힘들다.

따라서 기존 GAN의 loss function에 기초를 둔 두가지의 loss function을 사용했다

1. Edge-promoting adversarial loss ([Adversarial Loss](#AdversarialLoss))
2. A semantic content loss ([Content Loss](#ContentLoss))

Introduction

실제 사진을 수작업으로 만화 style 로 바꾸는 것은 매우 힘들다

최근 learning-base style transfer methods 가 상당한 이목을 끌었다. 특히 (GAN)

하지만 여러 성공에도 불구하고 최신 방법으로는 acceptable quality를 뽑아 내기에는 실패했다.

두가지 이유가 있는데 이는 위의 abstract에 나왔던 이유와 같다. ([Two Reason](#TwoReason))

Cartoon stylization의 목표는 이미지의 content를 유지하며 현실 사진의 manifold를 만화의 manifold로 변환하는 것이다.

이 논문에서는 훈련을 위해 unpaired image set를 사용하여 실제 사진에서 만화 image 로의 mapping을 효과적으로 학습하는 GAN 기반 접근법을 제시한다.

또한 두가지 효과적인 GAN에 기초한 효율적인 loss functions을 제시한다. [(Loss Function)](#LossFunction)

마지막으로 initialization phase를 추가한 더 효율적인 train method를 제시한다.

Related Work

1. NPR

* 많은 NPR 알고리즘들이 특정한 작가의 style을 모방하기 위해 개발되었다. 이러한 기술은 cel shading이라 불리며 game등을 만드는데 사용되어왔다. 하지만 현실의 사진이나 비디오를 만화로 바꾸는 데는 힘들었다. 주로 image filtering이나 formulations를 사용하지만 간단한 mathematical formulas를 사용하여 작가의 style을 capture 하기에는 무리가 있었다. 결과를 향상시키기 위해 segmentation of images/videos에 의존한 다른 방법이 제시되었지만 결과적으로 일반적인 images에 대처할 수 없었다.

1. Stylization with neural networks

* CNN 은 최근 상당한 관심을 받게 되었다. 전통적인 paired images를 필요로 하는 style transfer methods 와는 다르게 VGG network는 객체의 semantic features를 뽑아내는데 좋은 성능을 보였다. 결과 적으로 paired images를 필요로 하지 않는 더욱 강력한 style transfer methods가 개발되었다.
* 우선 CNN에 기반을 둔 neural style transfer (NST) method를 사용하였고 이들은 사전에 학습된 VGG 네트워크의 feature maps을 사용하여 컨텐츠를 표현하고 결과 이미지를 최적화하여 스타일 이미지의 content를 유지하면서 content image의 texture 정보를 일치시켰다. 이는 다양한 작가의 style을 변환하는데 좋은 결과를 보였으나 content 와 style 이미지가 서로 비슷해야 한다는 단점과 image가 여러 objects를 포함하고 있으면 의미상 다른 결과를 보였다. 또한 cartoon의 clear edges 와 smooth shading을 만들어내는데 실패하였다.
* LI and Wand는 CNNMRF를 사용하였지만 local matching이 오작동 할 수 있으며 정확하지 않은 결과를 도출했다.

1. Image synthesis with GANs

* GAN의 핵심 아이디어는 두가지 모델을 만들어 서로 학습시킨다는 것이다.
* cycleGAN은 unpaired data에 대한 학습이 가능하지만 동시에 두가지 GAN 모델을 학습시키는 것은 보통 시간이 매우 오래 걸린다. 또한 cycleGAN의 cartoonstylization의 결과는 그렇게 좋지 않다. 우리의 논문에서의 방법은 우리의 [Loss Function](#LossFunction) 덕분에 높은 quality의 cartoon images를 만들어 낼 수 있고, 학습을 더욱 효율적으로 할 수 있다.

1. Network architectures

* Deep neural networks를 학습하는데 notorious vanishing gradient problem때문에 어려움을 격을 수 있다. 최근 소개된 residual blocks는 training process를 줄이는데 매우 좋은 선택이 된다. 또한 이는 train 도중 발생하는 vanishing gradient 문제를 완화하도록 설계되었다.
* Training을 쉽게 하도록 도와주는 또다른 방법은 batch normalization이다. 이는 내부 공변량 변화에 대응하고 최소 점에 접근할 때 진동을 줄이도록 설계되었다.
* Leaky ReLu는 널리 사용되고 있는 activation function으로 gradient를 전달할 때 효율적이다.

CartoonGAN

GAN의 framework는 두가지 CNN으로 구성되어 있다. 이미지를 생성해 내는 generator G와 이미지를 판별하는 classifier인 discriminator D가 있다.

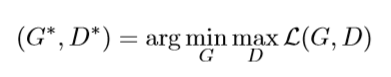
실제 사진의 manifold: P

Cartoon의 manifold: C

Train data: Sdata(p) = {pi |i = 1...N} ⊂ P and Sdata(c) = {ci |i = 1...M} ⊂ C (N과 M은 실제 사진과 cartoon image의 수이다.)

다른 GAN framework 와 마찬가지로 discriminator function D는 cartoon의 manifold를 다른 이미지사이에서 구별해 내고 G에대한 adversarial loss를 제공함으로써 G가 목표에 도달하도록 추진한다.

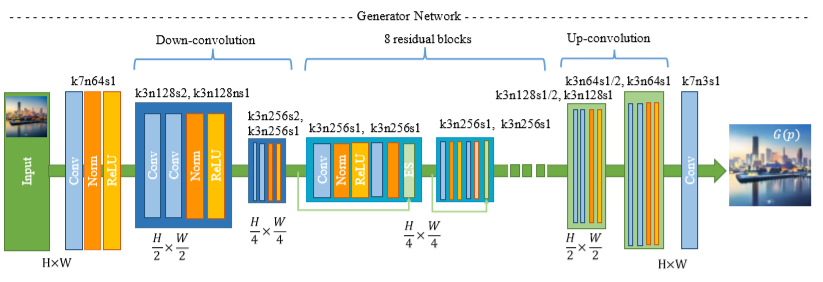
L을 loss function, G\*와 D\*를 각각 networks의 weights라 했을 때 다음과 같은 min-max problem을 해결하는 것이 목표가 된다.



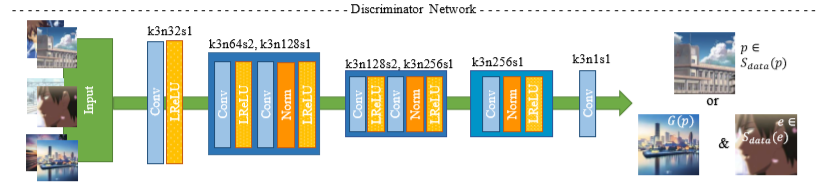
CartoonGAN architecture

Generator G는 flat convolution 단계부터 시작하여 이미지를 공간적으로 압축하고 인코딩하는 두개의 down convolution blocks으로 진행한다. 이때 downstream transformation을 위한 유용한 local signals이 추출된다??.... 이후 8개의 residual blocks가 content와 manifold feature을 구성한다.

마지막으로 두개의 up-convolution blocks를 거쳐 cartoon style의 images를 만들어 낸다.



Discriminator network D는 input image가 실제사진인지 cartoon image인지 판별한다. 다른 일반적인 full-image discriminator가 아닌 간단한 patch-level discriminator를 사용하였다. Flat layers 이후에 해상도를 낮추고 분류를 위한 필수 local 기능을 인코딩하기 위해 two-convolution layer block을 사용한다. 그 후, 특징 구성 블록 및 3\*3 convolutional layer가 분류 응답을 얻기 위해 사용된다.



Loss function

두가지 part로 구성되어 있다.

1. Adversarial loss (원하는 manifold 변환을 위한 generator network를 구동) :
2. Content loss (Cartoon stylization 도중 image content를 보존하기 위함) :

따라서

이때 w는 주어진 두가지 loss에 대한 weight이다. W가 커질수록 더 많은 content 정보가 보존된다. 이 논문에서는 w = 10으로 두었다.

1. Adversarial loss

* 양쪽 network G와 D에 적용되며, G가 cartoon transformation 하는데 영향을 끼친다.

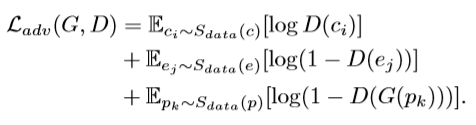
Output은 G의 output image가 cartoon image처럼 보이는 정도를 나타낸다.

일반적인 GAN frameworks처럼 간단한 학습으로는 좋은 결과를 내지 못한다는 것을 발견했다. 이는 clear edges가 cartoon의 중요한 특징이지만 전체 이미지에서 edge는 매우 작은 부분이기 때문이다. 따라서 clear edges가 없지만 shading은 정확한 결과는 discriminator에 혼동을 주게 된다.

이를 해결하기 위해 우리는 data를 추가하였다. 

이는 cartoon image에서 clear edges를 제거한 데이터다. Clear edges를 제거하는 방법으로 (1) Canny edge detector로 clear edges를 탐지하고 (2) edge 부분을 dilate 하며 (3) dilate 된 edge 부분에 Gaussian smoothing을 적용했다.

따라서 edge-promoting adversarial loss는 다음 과 같다.



1. Content loss

* Object preservation에 좋다고 입증된 미리 학습된 VGG network의 High-level feature maps를 사용한다. 따라서 content loss를 다음과 같이 정의한다.



l은 VGG layer의 특정한 feature maps를 나타낸다. 다른 image generation methods와는 다르게 input 사진과 만들어진 cartoon image 사이의 VGG feature maps의 l1 sparse regulation을 사용하였다. 이는 cartoon image가 실제 사진과는 다른 characteristics를 가지기 때문이다(I.e., clear edges, smooth shading). 적합한 VGG 레이어를 사용하더라도 거대한 Style 차이에 영향을 받을 수 있으며 이러한 차이는 지역 특성이 극적으로 변화는 지역에 집중되는 경향이 있다. l1 sparse가 l2 sparse에 비해 이러한 상황에 잘 대처하며 이는 cartoon style을 만드는데 매우 중요하다. 이 논문에서는 semantic content loss를 계산하는데 conv4\_4 layer의 feature maps를 사용하였다.

Initialization Phase

GAN 모델은 random initialization으로 매우 비선형 적이며 따라서 optimization이 최적이 아닌 local minimum으로 잡힐 수 있다. 이러한 수렴을 개선하기위해 새로운 initialization phase를 제시한다. 이 논문에서 Input image의 content를 재구성하는 데만 adversarial learning framework를 진행한다. 따라서 새로운 initialization phase에서는 content loss만을 이용하여 미리 학습된 generator network G를 사용한다. 이러한 간단한 initialization phase는 CartoonGAN의 조기 수렴 없이 빠른 구성으로 최적의 구성으로 수렴할 수 있게 한다.