

Generative Adversarial Text to Image Synthesis

Ha-larm

Text to Image

시각적 묘사 -> 이미지 생성

풀리지 않던 문제..

Zero shot caption-based retrieval

Sub Problem

1. 시각 특징 텍스트 표현 배우기
2. 이 특징들을 이용해 이미지 만들어내기

=> 이미 많은 발전을 이룬 분야들 !

Deep Learning으로 해결되지 않은 한가지 남겨진 문제점

=> 묘사를 정확하게 설명하는 그럴듯한 구성이 **매우 많다!**
(Image to Text도 같은 문제를 겪음..)

**순차적으로 분해될 수 있는 문자의 특성을
학습에 실용적으로 이용하자!**

Problem -> Conditional multi modality

Solution -> GAN + "smart" adaptive loss function!

Label 대신 텍스트 설명에 대한 모델 조건 !

Main contribution

- 1.Simple and effective GAN architecture
- 2.Training strategy that enables compelling text to image synthesis

GAN= G(generator)+D(discriminator)

Method

1.DC-GAN

2.conditioned on text features encoded by a hybrid character level convolutional recurrent neural net

Network architecture

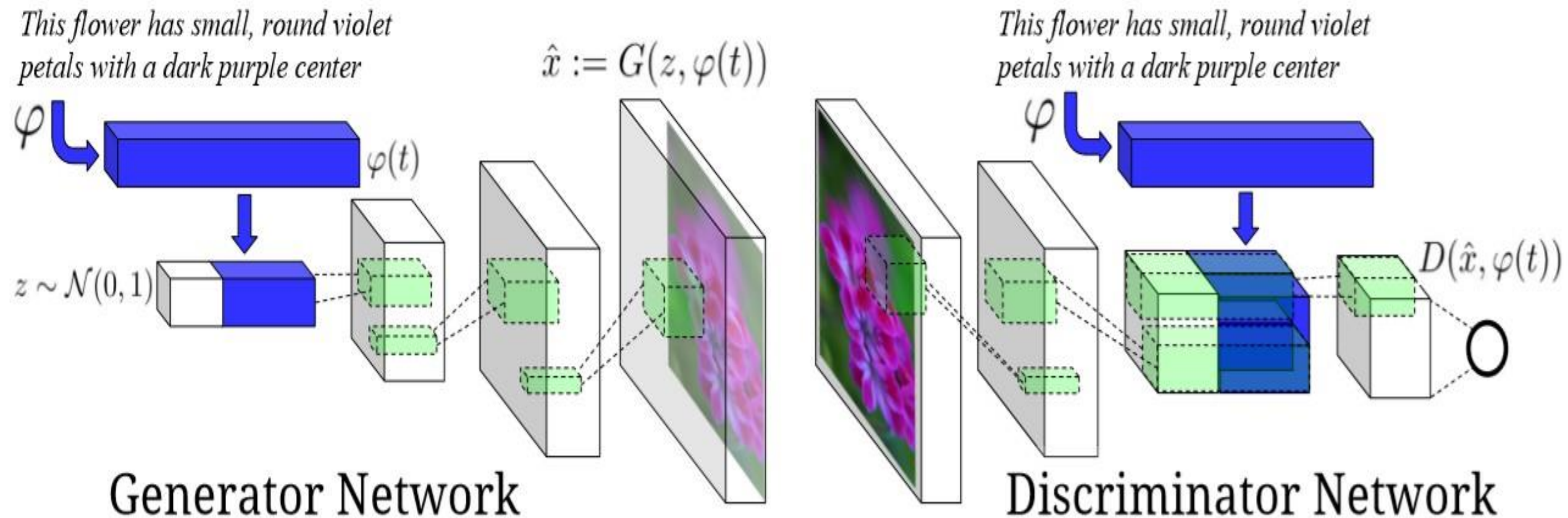


Figure 2. Our text-conditional convolutional GAN architecture. Text encoding $\varphi(t)$ is used by both generator and discriminator. It is projected to a lower-dimensions and depth concatenated with image feature maps for further stages of convolutional processing.

Generator

1. Sample from noise
2. Using text encoder(Using FC)
3. Leaky ReLU

Discriminator

1. 2 Stride convolution
2. Batch normalization
3. Leaky ReLU

학습의 시작에서 D가 조건을 무시하는 문제..

Naive GAN

Two kinds of input

- 1.Real image with matching text
- 2.Synthetic images with arbitrary text

Separate!

- 1.Unrealistic image(for any text)
- 2.Realistic images that mismatch

Modified GAN training algorithm to
separate these error source

Algorithm 1 GAN-CLS training algorithm with step size α , using minibatch SGD for simplicity.

- 1: **Input:** minibatch images x , matching text t , mis-matching \hat{t} , number of training batch steps S
 - 2: **for** $n = 1$ **to** S **do**
 - 3: $h \leftarrow \varphi(t)$ {Encode matching text description}
 - 4: $\hat{h} \leftarrow \varphi(\hat{t})$ {Encode mis-matching text description}
 - 5: $z \sim \mathcal{N}(0, 1)^Z$ {Draw sample of random noise}
 - 6: $\hat{x} \leftarrow G(z, h)$ {Forward through generator}
 - 7: $s_r \leftarrow D(x, h)$ {real image, right text}
 - 8: $s_w \leftarrow D(x, \hat{h})$ {real image, wrong text}
 - 9: $s_f \leftarrow D(\hat{x}, h)$ {fake image, right text}
 - 10: $\mathcal{L}_D \leftarrow \log(s_r) + (\log(1 - s_w) + \log(1 - s_f))/2$
 - 11: $D \leftarrow D - \alpha \partial \mathcal{L}_D / \partial D$ {Update discriminator}
 - 12: $\mathcal{L}_G \leftarrow \log(s_f)$
 - 13: $G \leftarrow G - \alpha \partial \mathcal{L}_G / \partial G$ {Update generator}
 - 14: **end for**
-

To Synthesize Realistic Images

- * 인코더는 내용물을 캡처한다.(contents)
- * 노이즈 샘플 z 는 배경을 캡처한다.(style)

To Achieve ConvNet Training(z)

$$\mathcal{L}_{style} = \mathbb{E}_{t, z \sim \mathcal{N}(0,1)} ||z - S(G(z, \varphi(t)))||_2^2$$



Figure 4. Zero-shot generated flower images using GAN, GAN-CLS, GAN-INT and GAN-INT-CLS. All variants generated plausible images. Although some shapes of test categories were not seen during training (e.g. columns 3 and 4), the color information is preserved.

Disentangling Style and content

스타일과 내용을 해체 함으로써 확장해보자!

Contents

Shape , Size , Color of each body part

Style

Other factor of variation

Text descriptions (content) Images (style)

The bird has a **yellow breast** with **grey** features and a small beak.

This is a large **white** bird with **black wings** and a **red head**.

A small bird with a **black head and wings** and features grey wings.

This bird has a **white breast**, brown and white coloring on its head and wings, and a thin pointy beak.

A small bird with **white base** and **black stripes** throughout its belly, head, and feathers.

A small sized bird that has a cream belly and a short pointed bill.

This bird is **completely red**.

This bird is **completely white**.

This is a **yellow** bird. The **wings are bright blue**.



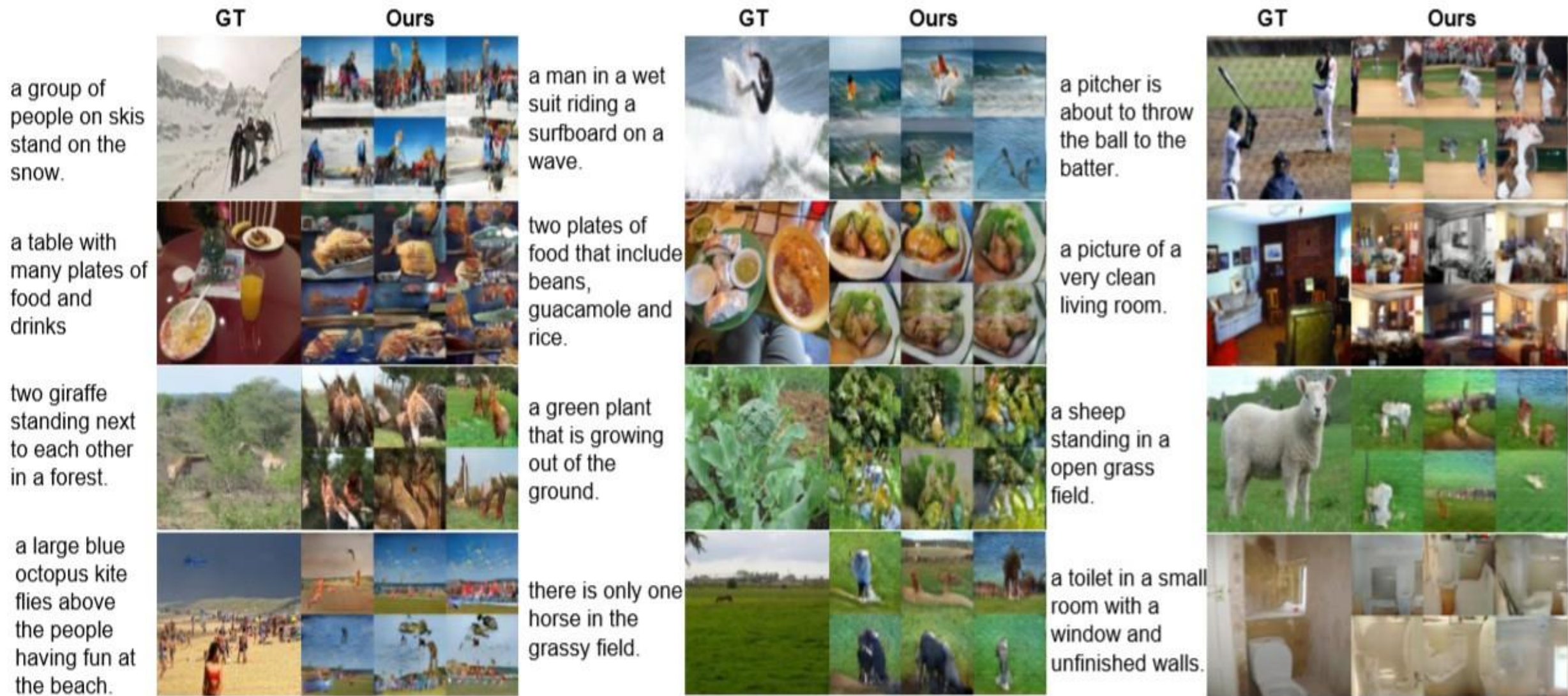


Figure 7. Generating images of general concepts using our GAN-CLS on the MS-COCO validation set. Unlike the case of CUB and Oxford-102, the network must (try to) handle multiple objects and diverse backgrounds.

Conclusion

Showed disentangling of Style and Content

Demonstrated the generalizability of generating images
With multiple object and variable background

Q & A

Thank you!

Ref paper

Generative Adversarial Text to Image Synthesis