Generative Adversarial Text-to-Image Synthesis Reed et al., 2016

Explainable Machine Learning Seminar

Frank Gabel - Thursday, 5 July

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results

◆□▶ ◆□▶ ◆目▶ ◆目▶ 目 のへぐ

Motivation

Who recognizes these celebrities?



Generative Adversarial Text-to-Image Synthesis

Motivation

Introductior

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results

▲□▶ ▲圖▶ ▲≣▶ ▲≣▶ ▲国 ● ● ●

Motivation

Who recognizes these celebrities?



Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results

You can't, they have been synthesized from

white noise.

Source : Progressive Growing of GANs for Improved Quality, Stability, and Variation (2017)

◆□▶ ▲□▶ ▲目▶ ▲□▶ ▲□▶

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models

Generative Adversarial Nets (GANs)

Conditional GANs

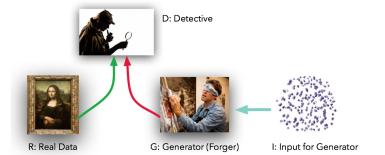
ション ふゆ ア キョン キョン ヨー もくの

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results

Discriminative models : p(Y|X) learn decision boundaries Generative models : p(X, Y) learn distributions

Introduction

Conceptually...



Mathematically...

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] +$$
(1)
$$\mathbb{E}_{x \sim p_{z}(z)} [\log(1 - D(G(z)))]$$

Generative Adversarial Text-to-Image Synthesis

Motivation

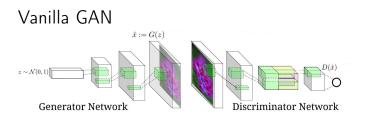
Introduction

Generative Models

Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results



 $G: \mathbb{R}^Z \to \mathbb{R}^{D \times D}$ and $D: \mathbb{R}^{D \times D} \to \{0, 1\}$

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

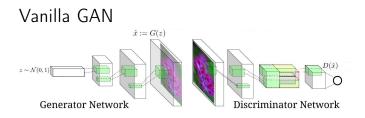
Generative Models

Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ・ うへぐ



$$G: \mathbb{R}^Z \to \mathbb{R}^{D \times D}$$
 and $D: \mathbb{R}^{D \times D} \to \{0, 1\}$

Is is possible to control the output of a GAN in a more meaningful way?

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models

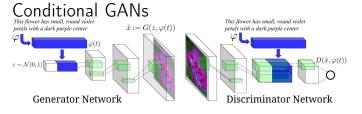
Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results

Vanilla GAN $\hat{x} := G(z)$ $z \sim \mathcal{N}(0,1)$ Generator Network $\hat{y} := G(z)$ $\hat{y} := G(z$

 $G: \mathbb{R}^Z \to \mathbb{R}^{D \times D}$ and $D: \mathbb{R}^{D \times D} \to \{0, 1\}$



 $G: \mathbb{R}^Z imes \mathbb{R}^T o \mathbb{R}^{D imes D}$ and $D: \mathbb{R}^{D imes D} imes \mathbb{R}^T o \{0, 1\}$

・ロト・日本・日本・日本・日本・日本

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

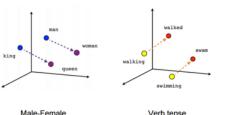
Conditional GANs

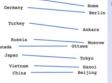
Architecture

Natural Language Processing Training Conditional GAN training dynamics Results Further Results

Introduction to Word Embeddings in NLP

- Map words to a high-dimensional vector space
- preserve semantic similarities :
 - president-power \approx prime minister
 - king-man+woman \approx queen.





Country-Capital

Generative Adversarial Text-to-Image Synthesis

Generative Models Adversarial Nets

Architecture

Natural Language Processing Training Conditional GAN training dynamics Further Results

Technique for embedding descriptions : deep symmetrical structural joint embedding (Reed et al., $2016) \Rightarrow$ idea is the same : preserve semantic similarities of sentences in the embedding space

Verb tense

Naive result

Generative Adversarial Text-to-Image Synthesis

Generative Models Adversarial Nets

Architecture Natural Language Processing

Training Conditional GAN training dynamics Results Further Results

an all black bird GT with a distinct thick, rounded bill.



this small bird has a vellow breast. brown crown, and black superciliary



a tiny bird, with a tiny beak, tarsus and feet, a blue crown, blue coverts, and black cheek patch





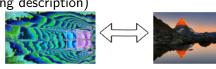
GAN

This does not work really well.

Conditional GAN training dynamics

Conditional GANs have an additional error source (unmatching description)

Naive GAN



Conditional GAN











A mountain scenery A dog and a cat at sunset cuddle.

Generative Adversarial Text-to-Image Synthesis

Generative Models Adversarial Nets

Architecture Natural Language Processing Training

Conditional GAN training dynamics

Results Further Results

▲ロ▶ ▲周▶ ▲ヨ▶ ▲ヨ▶ ヨー のへで

Matching-aware discriminator (GAN-CLS) so far :

- ▶ (real image, correct description) pairs \rightarrow classified as 1
- ▶ (fake image, correct description) pairs
 → classified as 0

ション ふゆ ア キョン キョン ヨー もくの

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training

Conditional GAN training dynamics

Matching-aware discriminator (GAN-CLS)

so far :

- (real image, correct description) pairs \rightarrow classified as 1
- ▶ (fake image, correct description) pairs
 → classified as 0

now :

 $\Rightarrow \mathsf{add}$

▶ (real image, false description) - pairs \rightarrow classified as 0

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

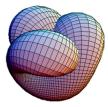
Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training

Conditional GAN training dynamics

Learning with manifold interpolation (GAN-INT) interpolate between embedding pairs



Build the arithmetic mean between embeddings from the training set : $t_{new} = \beta t_1 + (1 - \beta)t_2$ \rightarrow free training data Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training

Conditional GAN training dynamics

Inverting the generator for style transfer inject style information by training the *z* vector (background color, lighting etc.)

Train a style encoder network S : $s \leftarrow S(x), \ \hat{x} \leftarrow G(s, \phi(t))$

くしゃ 本面 そうせん ほう うめんろ

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training

Conditional GAN training dynamics

Matching-aware discriminator (GAN-CLS) introduce new types of samples Learning with manifold interpolation (GAN-INT) interpolate between embedding pairs Inverting the generator for style transfer inject style information (background color, pose etc.) Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training

Conditional GAN training dynamics

Results Further Results

◆□▶ ▲□▶ ▲目▶ ▲目▶ ▲□▶

Results - birds dataset

Generative Adversarial Text-to-Image Synthesis



Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics

Results Further Results

a tiny bird, with a this small hird has an all black bird tiny beak, tarsus and a vellow breast. GT with a distinct feet, a blue crown, brown crown, and thick, rounded bill. blue coverts, and black superciliary black cheek patch GAN GAN - CLS GAN - INT GAN - INT - CLS

Results - birds dataset - GAN-INT

'Blue bird with black beak' \rightarrow 'Red bird with black beak'



'Small blue bird with black wings' \rightarrow 'Small yellow bird with black wings'



'This bird is bright.' \rightarrow 'This bird is dark.'



Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics

Results Further Results

◆□▶ ▲□▶ ▲目▶ ▲目▶ ▲□▶

Results - more general dataset

GT

a group of people on skis stand on the snow.

a table with many plates of food and drinks

two giraffe standing next to each other in a forest.

a large blue octopus kite flies above the people having fun at the beach. Ours







a man in a wet suit riding a surfboard on a wave.

two plates of food that include beans, guacamole and rice.

a green plant that is growing out of the ground.

there is only one horse in the grassy field. GT



Ours

Generative Adversarial Text-to-Image Synthesis

Notivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics

Results - style transfer on birds dataset

Text descriptions Images (content) (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.





Generative Adversarial Text-to-Image Synthesis

Motivatior

Introduction

Generative Models Generative Adversarial Nets (GANs)

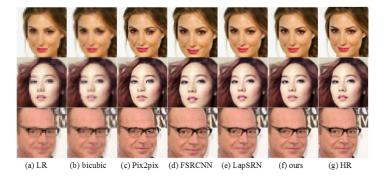
Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics

Results Further Results

・ロト ・日ト ・日ト ・日 ・ つへの

Recent application examples



Source : High-Quality Face Image Super-Resolution Using Conditional Generative Adversarial Networks (2017)

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

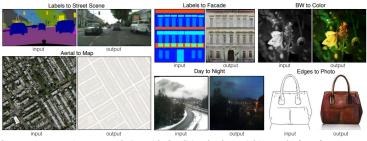
Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results Further Results

うせん 山 ふ 山 きょう かん うちょう

Generative Adversarial Text-to-Image Synthesis



Source : Image-to-Image Translation with Conditional Adversarial Networks (2017)

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results

Further Results

◆□▶ ◆□▶ ◆三▶ ◆三▶ ● ● ● ●

"In my opinion, among many interesting recent developments in deep learning, adversarial training is the most important one."

Yann LeCun, Director of Facebook AI

Generative Adversarial Text-to-Image Synthesis

Motivation

Introduction

Generative Models Generative Adversarial Nets (GANs)

Conditional GANs

Architecture Natural Language Processing Training Conditional GAN training dynamics Results

Further Results

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● ● ●