Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS COO

Er.

Generative Models

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neighbo

Evaluation

Discussio

Questions

References

Generating Natural Questions About an Image

[!

Ryan Callihan & Sarah Taylor

Seminar für Sprachwissenschaft Universität Tübingen

Janurary 12, 2018

Ryan Callihan & Sarah Taylor

Generating Questions fro Pictures

The Authors Objective

Image Recognition

Dataset

MS CO Bing

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

Evaluation

Discuss

References

Outline

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Datasets

MS COCO

Bing

Flicker

Generative Models

Maximum Entropy Language Model

Long Short-Term Memory

Gated Recurrent Neural Network

Retrieval Model

K-Nearest Neighbor

Evaluation

Discussion

Questions

References

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objectiv

Image Recognition using a CNN

Dataset

MS CO

D:

Flick

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

. ..

References

Getting Started

Make groups!

Ryan Callihan &

Generating
Questions from
Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Flicks

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mod

K-Nearest Neigh

Evaluation

Lvaiuatioi

Questions

References

Challenge One - Setup

Please go to: and enter in the code:

Ryan Callihan & Sarah Taylor

Generating Questions f

The Authors' Objective

Image Recognition

Datasets

MS COCO

Bing Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mod

K-Nearest Neighbor

Evaluation

Ouestions

References

Challenge One



Ryan Callihan & Sarah Taylor

Generating Questions fro Pictures

The Authors' Objectiv

Image Recognitionusing a CNN

Dataset

MS CO Bing

Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neight

Evaluation

Questions

References

Challenge One - Corpus Results



- ► Is this a religious ceremony?
- ► That looks very interesting, don't you think?
- ► What are they all gathered for?
- ▶ What are these people gathered for?
- ▶ Is this a satanic ritual?

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objection

Image Recognition using a CNN

Dataset

MS CO

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neigh

Evaluation

. ..

References

Challenge Two - Setup

Please go to: and enter in the code:

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objective

mage Recognition

Datasets

MS CO Bing

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

Evaluation

Discussion

References

Challenge Two



Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognitionusing a CNN

Dataset

Bing

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model K-Nearest Neighbor

Evaluation

Questions

References

Challenge Two - Corpus Results



- ► What are the people demonstrating about?
- What rally are they attending?
- ▶ What are these people protesting?
- ▶ What are they protesting?
- ▶ Who is in the gray jacket?

Ryan Callihan & Sarah Taylor

Generating Questions from

The Authors' Objective

Image Recognitio

Dataset

MS COO

Ring

Canada Madal

....

Long Short-Term Memor

Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

Lvaiuatio

D.5000

Questions

Reference

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objectiv

Image Recognition using a CNN

Dataset

MS CO

Ring

Elicke

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neig

Evaluation

Lvaluatio

. .

Reference

What is Image Recognition?

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

Image Recognitio

using a CNN

Dataset

MS COC

Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neighbor

Evaluation

Discussio

References

Image recognition and neural networks

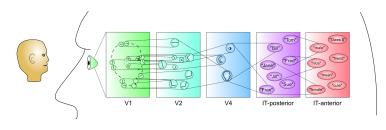


Figure: Visual network representation. Image from grey.colorado.edu [?]

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

Image Recognition

Dataset

MS COC

Bing Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

Evaluation

Evaluation

Questions

References

Convolutional neural network



Figure: CNN representation. Image from blog.floydhub.com/building-your-first-convnet [?]

Generating Natural

Ryan Callihan &

References

CNN visualized

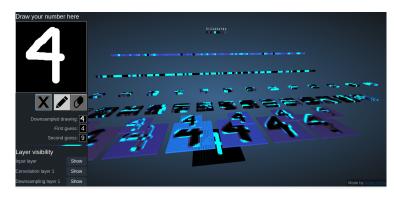


Figure: CNN visualized. Image from scs.ryerson.ca/ [?]

Generating Natural

Ryan Callihan &

Evaluation

References

Why Does It Work Well for Images?

Ryan Callihan & Sarah Taylor

Generating

The Authors' Objective

Image Recognitio

S . .

MS COO

Bing

Generative Model

Maximum Entrop

Long Short-Term Memor

Gated Recurrent Neural

Retrieval Mode

K-Negrest Neigh

Evaluation

D .500a.

.

Reference

Generating Natural

Ryan Callihan &

Ryan Callihan & Sarah Taylor

Generating

The Authors' Objectiv

Image Recognitio

D.

... ...

Bing

Generative Models

Maximum Entron

Long Short-Term Memor

Gated Recurrent Neural

Retrieval Mode

K-Negreet Neigh

Evaluation

Lvaluation

. .

Reference

Ryan Callihan & Sarah Taylor

Generating

The Authors' Objective

Image Recognitio

Dataset

MS COC

Bing

Generative Model

Maximum Entrop

Long Short-Term Memor

Gated Recurrent Neural

Retrieval Mode

K-Negrest Neigh

Evaluation

Lvaiuatioi

. .

Reference

Ryan Callihan & Sarah Taylor

Questions from Pictures

Image Recognition

using a CNN

Dataset

Bing Flicker

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model
K-Nearest Neighbor

Evaluation

Ouestions

References

Generative Models



Caption Bot [?]

captionbot.ai was used throughout this paper to automatically generate captions. It is a Microsoft project based on the Computer Vision API, Emotion API, and Bing Image Search API.

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognition

Dataset

MS COC

Bing Flicker

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neighbo

Evaluation

Discussio

Questions

References

Caption Bot example



I think it's a person smiling for the camera and she seems 🗟.



Ryan Callihan & Sarah Taylor

Generating
Questions from
Pictures
The Authors' Objective

Image Recognition using a CNN

Dataset

Bing Flicker

Generative Mode

Maximum Entropy Language Model Long Short-Term Memor Gated Recurrent Neural Network

Retrieval Mode K-Nearest Neighbor

Evaluatio

Questions

References

Computer Vision API example



FEATURE NAME:	VALUE
Description	("tags": ["person", "man", "indoor", "table", "sitting", "holding", "food", "woman", "glasses", "people", "posing", "drinking", "wine", "restraumer", 'plate", "siming", 'pizza", "phone", "young, "standing", "store", "group", "white"]. "captions": ["I "text": "a man sitting at a table in a restaurant", "confidence": 0.9105153].)
Tags	[{"name": "person", "confidence": 0.999498367), ("name": "man", "confidence": 0.926230047), ("name": "indoor", "confidence": 0.8648256), ("name": "restaurant", "confidence": 0.193121776}]
Image format	"Jpeg"

- ► Description "tags": ["person", "man", "indoor", "table",
 "sitting", "holding", "food", "woman", "glasses", "people",
 "posing", "drinking", "wine", "restaurant", "plate", "smiling",
 "pizza", "phone", "young", "standing", "store", "group", "white"
], "captions": ["text": "a man sitting at a table in a
 restaurant", "confidence": 0.9105153]
- Faces ["age": 25, "gender": "Male", "faceRectangle": "top": 94, "left": 149, "width": 79, "height": 79, "age": 33, "gender": "Male", "faceRectangle": "top": 97, "left": 343, "width": 79, "height": 79]

an Image

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objective

Image Recognition using a CNN

Dataset

MS COO

Bing

Flick

Generative Model

Maximum Entrop

Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

. .

Reference

MELM

LSTM

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objectiv

Image Recognition using a CNN

Dataset

MS CO

D....8

Generative Model

Maximum Entropy

Long Short-Term Memor

Retrieval Mode

K-Nearest Neight

Evaluation

. .

Reference

Keierenc

Ryan Callihan & Sarah Taylor

Saran Taylor

Questions from Pictures

The Authors' Objectiv

Image Recognition using a CNN

Dataset

MS COO

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

. ..

Reference

GRNN

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objective

Image Recognition using a CNN

Datacet

MS CO

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Patrioval Made

K-Nearest Neigh

Evaluation

D .500a.

.

Reference

Retrieval Model

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objectiv

Image Recognition using a CNN

D.

MS COO

D:

Elick

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neig

Evaluation

Evaluation

. .

Questions

Reference

K-Nearest Neighbor

BLEU

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objectiv

Image Recognition

Dataset

MS CO

D....8

Generative Model

Maximum Entropy Language Model Long Short-Term Memory

Date Const. March

Evaluation

Evaluation

D.50005

D-f----

Reference

Ryan Callihan & Sarah Taylor

Generating Questions from

The Authors' Objective

Image Recognition using a CNN

Datacet

MS CO

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Patrioval Made

IZ NI . NI S I

Electrical Control

Evaluation

D 1500551

Reference

Evaluation

Generating Natural

Ryan Callihan &

References

BLEU & METEOR

Generating Natural

Ryan Callihan &

Evaluation

References

Discussion - Authors' Thoughts

Ryan Callihan & Sarah Taylor

Generating Questions from

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory

Retrieval Mode

K-Nearest Neigh

Evaluation

Outstiene

Reference

Questions

Ryan Callihan & Sarah Taylor

Generating Questions from

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Generative Mode

Maximum Entropy Language Model

Long Short-Term Memor Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

. .

Reference

References I