

[captionbot.ai](#)

Apps

Getting Started

Imported From Fire...

Bookmarks



Other bookmarks

CaptionBot



I think it's a man flying through the air while riding a skateboard.



onlybhawnagupta@gmail.com



How did I do?





I think it's a group of people posing for the camera.

onlybhawnaagupta@gmail.com



How did I do?



CaptionBot



I think it's a train crossing a bridge over a river.



How did I do?

Image Captioning with keras

Teaching Computers to describe pictures



Harshall Lamba [Follow](#)

Nov 4, 2018 · 21 min read

Table of Contents:

1. Introduction

onlybhawnagupta@gmail.com

2. Motivation

3. Prerequisites

4. Data collection

5. Understanding the data

6. Data Cleaning

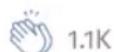
7. Loading the training set

8. Data Preprocessing — Images

Towards Data Science

Sharing concepts,
ideas, and codes.

Following ▾



1.1K



- Self driving cars — Automatic driving is one of the biggest challenges and if we can properly caption the scene around the car, it can give a boost to the self driving system.

onlybhawnagupta@gmail.com

- Aid to the blind — We can create a product for the blind which will guide them travelling on the roads without the support of anyone else. We can do this by first converting the scene into text and then the text to voice. Both are now famous applications of Deep Learning. Refer this [link](#) where its shown how Nvidia research is trying to create such a product.

- CCTV cameras are everywhere today, but along with viewing the world, if we can also generate relevant captions, then we can raise alarms as soon as there is some malicious activity going on somewhere. This could probably help reduce some crime and/or accidents.

- Automatic Captioning can help, make Google Image Search as good as Google Search, as then every image could be first converted into a

Dataset

Flickr Image dataset

Flickr Image captioning dataset

onlybhawnaagupta@gmail.com

Hsankesara • updated a year ago (Version 1)

Data Kernels (12) Discussion (1) Activity Metadata Download (4 GB) New Notebook

Usability 7.5

License CCO: Public Domain

Tags nlp, image data, image processing

Description

The Flickr30k dataset has become a [standard benchmark for sentence-based image description](#). This paper presents Flickr30k Entities, which augments the 158k captions from Flickr30k with 244k coreference chains, linking mentions of the same entities across different captions for the same image, and associating them with 276k manually annotated bounding boxes. Such annotations are essential for continued progress in automatic image description and grounded language understanding. They enable us to define a new benchmark for localization of textual entity mentions in an image. We present a strong baseline for this task that combines an image-text embedding, detectors for common objects, a color classifier, and a bias towards selecting larger objects. While our baseline rivals in accuracy more complex state-of-the-art models, we show that its gains cannot be easily parlayed into improvements on such tasks as image-sentence retrieval, thus underlining the limitations of current methods and the need for further research.

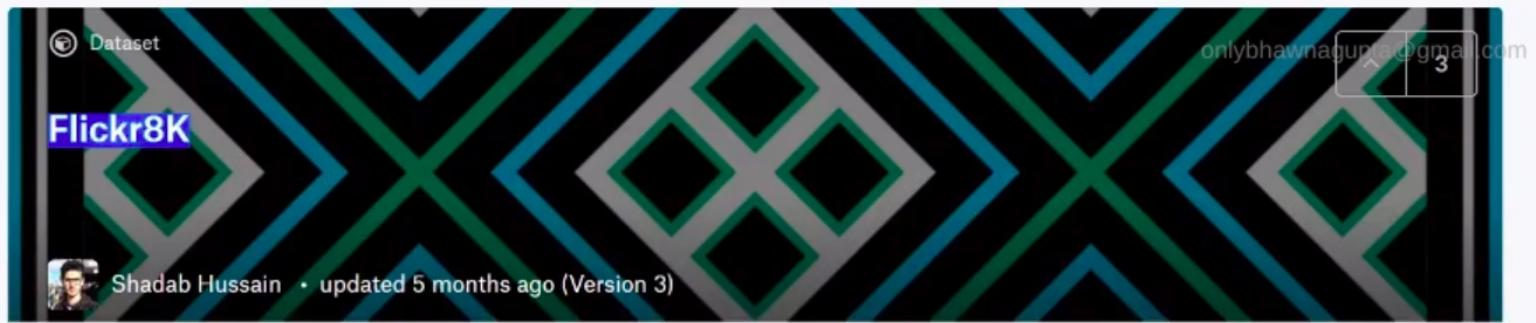
Flickr8K | Kaggle | Flickr Image dataset | Kaggle | prateek27/Image-Capti... | Image Captioning with... | CaptionBot - For pictur... | Downloads

kaggle.com/shadabhussain/flickr8k/downloads/flickr8k.zip/3

Apps Getting Started Imported From Fire... Bookmarks

Other bookmarks

kaggle Search



Your Dataset download has started.
Show your appreciation with an upvote

3



Usability 1.9

Tags No tags yet

Data (1 GB)



Data Sources

model_weights.h5

train_encoded_images.p

About this file

No description yet

Image Captioning (Project)



"A cute little dog
sitting in grass"

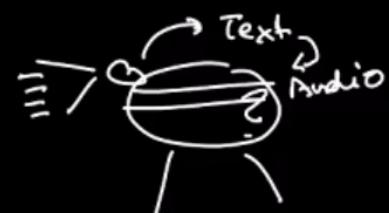
onlybhawnagupta@gmail.com

↑
input
(image) Goal:

↑
output
(sentence)

Applications

① Visual Aid for Blind



Flick8K

8000 image

5 captions

for each image

② Google Search



Text Query

③ Automatic Surveillance (CCTV Camera's)

8000 images
5 captions
for each img ③

Automatic Surveillance (CCTV Camera's)

onlybhawnagupta@gmail.com

X Y
[img sent₁, sent₂, ... sent₅]

Flick8K
Dataset

must do
→ {MLP, CNN, RNN, Language Model,
word Embeddings, Transfer Learning,
Pre-processing}

$$q = f(x)$$