

Image Captioning(Final-term Project)

Image Captioning

• Captions of an image



000000184613.jpg

- A boy holding an umbrella while standing next to livestock.
- A child holding a flowered umbrella and petting a yak.
- A young man holding an umbrella next to a herd of cattle.
- a young boy barefoot holding an umbrella touching the horn of a cow
- A young boy with an umbrella who is touching the horn of a cow

Goal

You have to train your model with some captions above then generate another caption for other image.

Dataset

• json file

```
Data.json
{
  'images': [
    {
      'file_name' : '000000184613.jpg',
      'height': 336,
      'width': 500,
      'image_id': 184613,
      'captions': ['A boy holding an umbrella while standing next to
livestock.', 'A child holding a flowered umbrella and petting a yak.', 'A young
man holding an umbrella next to a herd of cattle.', 'a young boy barefoot
holding an umbrella touching the horn of a cow', 'A young boy with an umbrella
who is touching the horn of a cow']
    }, . . . . .
  ]
}
```

Dataset

Use torch.utils.data.Dataset

Make pair of (image, caption)

You can use all of captions, or just one (your choice)

Testing the model

We will test your project with undisclosed dataset. You must generate a caption to corresponding each image.

We recommend you to divide the images into train/valid/test and test yourself.

Data.zip

• Upload data to Colab.

1. Download “data.zip” on Icampus.

2. Upload “data.zip” on you Google drive.

- You can use unlimited Google drive using school mail
(YOUR_ID@g.skku.edu)

3. Mount the drive.

```
from google.colab import drive  
drive.mount("/content/drive")
```

(click the URL to get your authorization code.)

4. Copy the file path and unzip using
`!unzip “path”`



Restrictions

● Restrictions train.py

1. train.py

- If you want to use pre trained CNN model, you can use only torchvision.models.
- If you want to use pre trained RNN model, you can use pre-trained model that you make by yourself.
- Or you can use pre-trained char/word embedding.
- Train and save the model as “[NAME_Student Id].pt”

Restrictions

● Restrictions test.py

2. test.py

- Load the model that saved at train.py.
- When you run test.py, generate a caption for all images in test folder named 'test' which is in the same directory. And make a json file(NAME_result.json) described as follow at the same directory.

```
{
  'images': [
    {
      'file_name' : '000000184613.jpg',
      'captions': 'A boy holding an umbrella while standing next to
livestock.'
    }, . . . . .
  ]
}
```

Report

● Report(How you implemented / How you tried.)

1. How you implemented

- A full description of each process you have taken to solve the problem(preprocessing, model construction, etc.)
- How you train and test the model and the result of the test.

2. What have you tried.

- All the things you studied to solve the problem.
- The problem may be difficult to achieve high performance. But write down what you've studied on your own to get the performance and to solve the problems.

3. Restrictions

- Use ms word with 11pt & maximum 3 pages.

Submission

• [Student Id].zip

Codes

- Train file → `'train_[NAME_Student Id].py'`
- Test file → `'test_[NAME_Student Id].py'`
- Any other python files should be under folder `'utils_[NAME_Student Id]'`

Model file → `'model_[NAME_Student Id].pt'`


Report → `'[NAME_Student Id].pdf'` (English / Korean OK)


Put all files in same directory named `'[NAME_Student ID]'` and **zip this folder** → `'[NAME_Student Id].zip'`






Submission due date **19 December 2020** and **No Late Submission.**

Submission

Submit this file.

이름	수정한 날짜	유형
 홍길동_2020123456	2020-12-07 오후 11:04	압축(ZIP) 파일

이름	수정한 날짜	유형	크기
 홍길동_2020123456	2020-12-07 오후 11:04	파일 폴더	

이름	수정한 날짜	유형	크기
 utils_홍길동_2020123456	2020-12-07 오후 11:03	파일 폴더	
 model_홍길동_2020123456.pt	2020-12-07 오후 11:03	PT 파일	0KB
 test_홍길동_2020123456	2020-12-07 오후 11:02	Python File	1KB
 train_홍길동_2020123456	2020-12-07 오후 11:02	Python File	1KB
 홍길동_2020123456	2020-12-07 오후 11:03	Microsoft Edge P...	356KB

Criteria

1. Model performance 50%

- accuracy score
- similarity score

2. Report 50%

- how well do you study..
- how well do you analyze..
- ...