St.Peter's College Of Engineering And Technology

Avadi, Chennai-54

DEPARTMENT OF INFORMATION TECHNOLOGY

NAAN MUDHALVAN PROJECT REPORT

IMAGE TO TEXT

Team Members:

S.SHIVANI

B.Tech IT

112721205015

IMAGE TO TEXT

INDEX

Abstract:

Image captioning is the task of describing the content of the image using textual representation. It has been used in many applications such as semantic tagging, image retrieval, early childhood learning, human-like robot–robot interactions, visual question answering tasks, and medical diagnosis. In the medical field, automatic captioning assists medical professional in diagnosis, disease treatment, and follow-up recommendations. Many efforts have been put forward to develop accurate machine learning algorithms for medical image captioning but systems still provide low-quality descriptions. Recently, particular attention has been focused on providing robust explainability modules for many machine learning tasks.

INTRODUCTION:

In general, deep learning based captioning relies on encoder–decoder models which are black boxes of two components cooperating to generate new captions for images. Image captioning builds a bridge between natural language and image processing, making it difficult to understand the correspondence between visual and semantic features. We present, in this chapter, an explainable approach that provides a sound interpretation of the attention-based encoder–decoder model for image captioning. It provides a visual link between the region of medical image and the corresponding wording in the generated sentence. For that, a self-attention mechanism is employed to compute word importance for the semantic features, and attention mechanism is used to compute the most relevant regions of the image considered by the model to generate the caption sequences. We evaluate the performance of the model and provide samples from the ImageCLEF medical captioning dataset.

Retrieval-based method, which relies on an image search engine to procure visual words based on features from images.

Template-based method, which is a two-staged strategy through object detection and classification, as well as sentence generation based on an object's attribute and the relationship between objects and environments

PROGRAM:

```
pip install gradio
pip install TensorFlow
pip install requests
pip install transformers
import gradio as gr
import requests
from PIL import Image
from transformers import AutoProcessor, AutoModelForVision2Seq

model = AutoModelForVision2Seq.from_pretrained("microsoft/kosmos-2-
patch14-224")
processor = AutoProcessor.from_pretrained("microsoft/kosmos-2-patch14-
224")

def generate_description(image_url):
    try:
        image = Image.open(requests.get(image_url, stream=True).raw)
        image.save("new_image.jpg")
        image = Image.open("new_image.jpg")
        prompt = "<grounding>An image of"
        inputs = processor(text=prompt, images=image,
return_tensors="pt")

        generated_ids = model.generate(
            pixel_values=inputs["pixel_values"],
            input_ids=inputs["input_ids"],
            attention_mask=inputs["attention_mask"],
            image_embeds=None,

image_embeds_position_mask=inputs["image_embeds_position_mask"],
            use_cache=True,
            max_new_tokens=128,
        )
        generated_text = processor.batch_decode(generated_ids,
skip_special_tokens=True)[0]

        processed_text, _ =
processor.post_process_generation(generated_text)
        return processed_text
    except Exception as e:
        return str(e)
```

```
iface = gr.Interface(
    fn=generate_description,
    inputs="text",
    outputs="text",
    title="Image Description Generator",
    description="Enter the URL of an image to generate a description.",
    examples=[
        ["https://buffer.com/cdn-cgi/image/w=1000,fit=contain,q=90,f=auto/library/content/images/size/w600/2023/10/free-images.jpg"]
    ]
)
iface.launch()
```

NEW IMG .JPG:

OUTPUT:

AN INTERFACE WILL BE CREATED

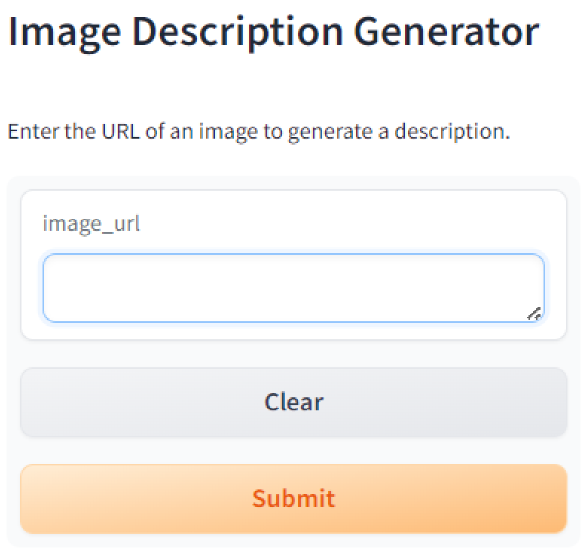
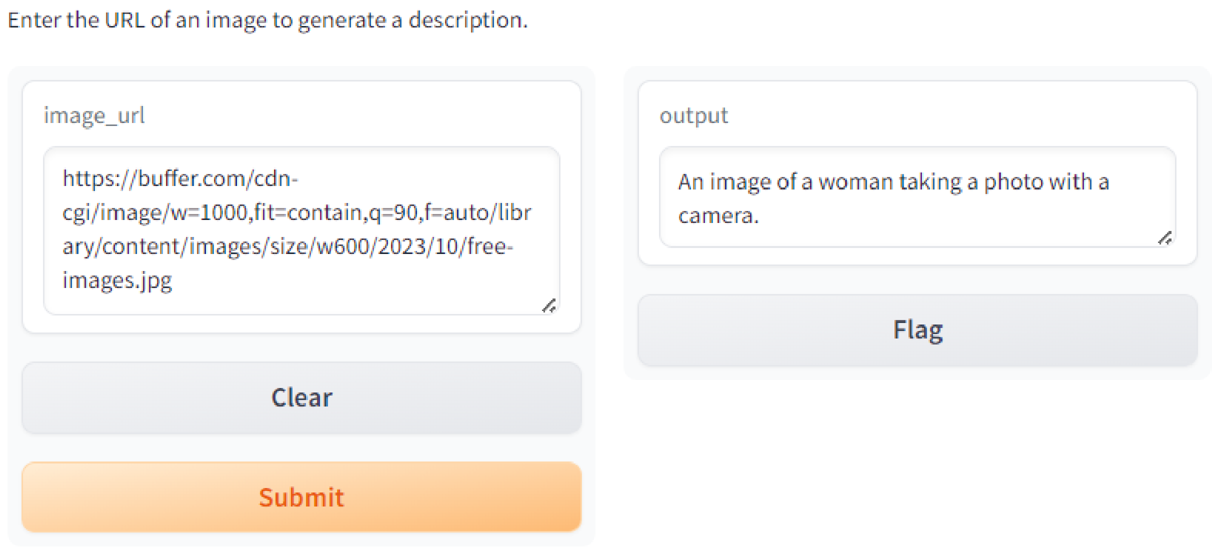


Image Description Generator

Enter the URL of an image to generate a description.

AN OUTPUT WILL BE GENERATED



Enter the URL of an image to generate a description.

image_url: https://buffer.com/cdn-cgi/image/w=1000,fit=contain,q=90,f=auto/library/content/images/size/w600/2023/10/free-images.jpg

output: An image of a woman taking a photo with a camera.

RESULT:

In conclusion, image captioning represents a powerful fusion of computer vision and natural language processing, bridging the gap between visual content and textual understanding. The article has elaborated on the significance of image captioning and its impact across various domains.