

Narzędzie do generowania tekstu z obrazków przy użyciu transformatorów 'image-to-text' z portalu Hugging Face

Biegan Maciej

Adamowicz Jan

Bobak Jan

Dziedzic Kamil

Technologia image-to-text

- Generowanie opisów tekstowych na podstawie obrazków
- Digitalizacja
- Filtrowanie zdjęć w social mediach
- Wspieranie dostępności dla osób z niepełnosprawnościami



Platforma Hugging Face

- Szeroki wybór gotowych modeli
- Modele przeszkolone na dużych zbiorach danych



Wybrane modele i ich cechy

- Modele z rodziny BLIP (Bootstrapped Language-Image Pre-training)
- Model Salesforce/blip-image-captioning-large
- Model Salesforce/blip-image-captioning-base
- jedne z najpopularniejszych i najlepiej ocenianych modeli

Models 723

Filter by name

 Salesforce/blip-image-captioning-large
Image-to-Text • Updated Dec 7, 2023 • 1.97M • ⚡ • ❤️ 1.2k

 microsoft/trocr-base-handwritten
Image-to-Text • Updated May 27 • 853k • ❤️ 347

 Salesforce/blip-image-captioning-base
Image-to-Text • Updated Aug 1, 2023 • 1.78M • ⚡ • ❤️ 524

Działanie modelu BLIP

- Koder obrazu
- Dekoder tekstu
- Proces przetwarzania:
 1. Wstępne przetwarzanie obrazu
 2. Kodowanie obrazu
 3. Generowanie tekstu



Implementacja rozwiązania

Funkcja `initialize_model`

Cel: Inicjalizuje procesor i model BLIP do generowania opisów obrazów.

Działanie:

- ❑ Tworzy katalog cache dla danych Hugging Face.
- ❑ Ładuje:
 - ❑ **Procesor:** Przygotowuje dane wejściowe (np. obrazy) w formacie zgodnym z modelem.
 - ❑ **Model:** Generuje opisy na podstawie przetworzonych danych.
- ❑ Obsługuje błędy (zwraca None w przypadku niepowodzenia).

```
def initialize_model():  
    try:  
        cache_dir = Path.home() / ".cache" / "huggingface"  
        cache_dir.mkdir(parents=True, exist_ok=True)  
  
        processor = BlipProcessor.from_pretrained(  
            "Salesforce/blip-image-captioning-base",  
            cache_dir=cache_dir,  
            local_files_only=False,  
        )  
        model = BlipForConditionalGeneration.from_pretrained(  
            "Salesforce/blip-image-captioning-base",  
            cache_dir=cache_dir,  
            local_files_only=False,  
        )  
        return processor, model  
    except Exception as e:  
        print(f"Error initializing model: {e}")  
        return None, None
```


Implementacja rozwiązania

Funkcja: *generate_text_from_image*

Służy do generowania tekstowych opisów obrazów przy użyciu modelu BLIP. Obsługuje obrazy lokalne i z URL. Funkcja przetwarza obraz, generuje tensor wejściowy, stosuje odpowiednie parametry i zwraca wygenerowany opis oraz czas przetwarzania.

Parametry generowania (params):

☐ **simple:**

- ☐ max_length=100: Maksymalna długość opisu.
- ☐ min_length=10: Minimalna długość opisu.
- ☐ num_beams=5: Liczba ścieżek poszukiwań w beam search.
- ☐ early_stopping=True: Zatrzymanie generacji po znalezieniu dobrego wyniku.
- ☐ no_repeat_ngram_size=3: Zapobieganie powtórzeniom 3-gramów.

☐ **advanced:**

- ☐ max_length=300: Dłuższe opisy.
- ☐ min_length=30: Większa szczegółowość.
- ☐ num_beams=30: Dokładniejsze przeszukiwanie wyników.
- ☐ early_stopping=False: Kontynuacja generacji do pełnej długości.
- ☐ no_repeat_ngram_size=5: Mniejsze powtórzenia.

```
def generate_text_from_image(image_path, mode="simple",
                             processor=None, model=None):
    if processor is None or model is None:
        return "Model not initialized properly", 0

    try:
        start_time = time.time()

        if image_path.startswith("http"):
            response = requests.get(image_path, stream=True)
            image = Image.open(response.raw).convert("RGB")
        else:
            image = Image.open(image_path).convert("RGB")

        inputs = processor(images=image, return_tensors="pt")

        params = {
            "simple": {
                "max_length": 100,
                "min_length": 10,
                "num_beams": 5,
                "early_stopping": True,
                "no_repeat_ngram_size": 3
            },
            "advanced": {
                "max_length": 300,
                "min_length": 30,
                "num_beams": 30,
                "early_stopping": False,
                "no_repeat_ngram_size": 5
            }
        }

        out = model.generate(**inputs, **params[mode])

        description = processor.decode(out[0], skip_special_tokens=True)
        generation_time = time.time() - start_time

        return description, generation_time
    except Exception as e:
        return f"Error processing image: {e}", 0
```

Przykłady testów

Model Salesforce/blip-image-captioning-base (0.9gb)



```
• (myenv) kamildziedzic@MacBook-Pro-Kamil pjn % python3 projekt.py
```

```
Menu:
```

1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź

```
Wybierz opcję: 1
```

```
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/467186778_2013511619077349_1452096888790434597_n.png
```

```
Opis (simple): a dog sitting on top of a tree branch
```

```
Czas generowania: 2.17 sekund
```

```
Menu:
```

1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź

```
Wybierz opcję: 2
```

```
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/467186778_2013511619077349_1452096888790434597_n.png
```

```
Opis (advanced): a dog sitting on top of a tree branch in the middle of the image is a golden retrieve sitting on a log in the dog is looking up at the camera
```

```
Czas generowania: 16.53 sekund
```

Model Salesforce/blip-image-captioning-large (1.8gb)

```
Menu:
```

1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź

```
Wybierz opcję: 1
```

```
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/467186778_2013511619077349_1452096888790434597_n.png
```

```
Opis (simple): there is a dog that is sitting on a branch in the grass
```

```
Czas generowania: 3.35 sekund
```

```
Menu:
```

1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź

```
Wybierz opcję: 2
```

```
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/467186778_2013511619077349_1452096888790434597_n.png
```

```
Opis (advanced): there is a dog sitting on a branch in the middle of a field with trees in the background and a dog sitting on the branch in the foreground of the foreground
```

```
Czas generowania: 21.18 sekund
```


Przykłady testów

Model Salesforce/blip-image-captioning-base (0.9gb)



```
○ (myenv) kamildziedzic@MacBook-Pro-Kamil pjn % python3 projekt.py

Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 1
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/JanMatejkoRejtan.jpg

Opis (simple): a painting of a group of people in a room
Czas generowania: 1.82 sekund

Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 2
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/JanMatejkoRejtan.jpg

Opis (advanced): a painting of a group of people sitting in front of a painting of a man in a white suit and a woman in a red dress, a woman in a white dress, and a man in a man in a black suit and a white dress
Czas generowania: 24.09 sekund
```

Model Salesforce/blip-image-captioning-large (1.8gb)

```
○ (myenv) kamildziedzic@MacBook-Pro-Kamil pjn % python3 projekt.py

Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 1
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/JanMatejkoRejtan.jpg

Opis (simple): there is a painting of a large group of people in a room
Czas generowania: 5.20 sekund

Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 2
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/JanMatejkoRejtan.jpg

Opis (advanced): this is a painting of a group of people standing in front of a painting of a man in a suit and a woman sitting on a chair in front of another man ' s lap
Czas generowania: 17.54 sekund
```

Przykłady testów

Model Salesforce/blip-image-captioning-base (0.9gb)

```
Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 1
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/IMG_4557.jpeg

Opis (simple): three men holding coffee mugs and smiling at the camera
Czas generowania: 3.02 sekund
```

```
Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 2
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/IMG_4557.jpeg
```

```
Opis (advanced): a group of three men standing next to each other two are holding coffee mugs and one is holding a coffee mug and the other two are holding mugs
Czas generowania: 16.52 sekund
```



Model Salesforce/blip-image-captioning-large (1.8gb)

```
Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 1
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/IMG_4557.jpeg
```

```
Opis (simple): there are three men standing next to each other holding coffee cups
Czas generowania: 1.90 sekund
```

```
Menu:
1. Generuj opis z URL obrazu (simple)
2. Generuj opis z URL obrazu (advanced)
3. Wyjdź
Wybierz opcję: 2
Wprowadź URL obrazu: /Users/kamildziedzic/Downloads/IMG_4557.jpeg
```

```
Opis (advanced): there are three men standing next to each other holding coffee mugs in front of a sign that says dziekkanat on the side of the wall
Czas generowania: 22.64 sekund
```




REVUELTO

Dziękujemy za uwagę

(teraz live show naszego rozwiązania na żywo)