

# deep-learning-project

April 20, 2024

```
[ ]: !pip install transformers timm einops -q
```

```
[ ]: !pip install flash-attn --no-build-isolation
```

Collecting flash-attn

Downloading flash\_attn-2.5.7.tar.gz (2.5 MB)

2.5/2.5 MB

9.3 MB/s eta 0:00:00

Preparing metadata (setup.py) ... done

Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages (from flash-attn) (2.2.1+cu121)

Requirement already satisfied: einops in /usr/local/lib/python3.10/dist-packages (from flash-attn) (0.7.0)

Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from flash-attn) (24.0)

Collecting ninja (from flash-attn)

Downloading ninja-1.11.1.1-py2.py3-none-manylinux1\_x86\_64.manylinux\_2\_5\_x86\_64.whl (307 kB)

307.2/307.2

kB 30.8 MB/s eta 0:00:00

Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (3.13.4)

Requirement already satisfied: typing-extensions>=4.8.0 in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (4.11.0)

Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (1.12)

Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (3.3)

Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (3.1.3)

Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (2023.6.0)

Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (12.1.105)

Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (12.1.105)

Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in

```

/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (12.1.105)
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (8.9.2.26)
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (12.1.3.1)
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (11.0.2.54)
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (10.3.2.106)
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (11.4.5.107)
Requirement already satisfied: nvidia-cuspars-cu12==12.1.0.106 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (12.1.0.106)
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (2.19.3)
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from torch->flash-attn) (12.1.105)
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-
packages (from torch->flash-attn) (2.2.0)
Requirement already satisfied: nvidia-nvjitlink-cu12 in
/usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-
cu12==11.4.5.107->torch->flash-attn) (12.4.127)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2->torch->flash-attn) (2.1.5)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-
packages (from sympy->torch->flash-attn) (1.3.0)
Building wheels for collected packages: flash-attn
  Building wheel for flash-attn (setup.py) ... done
  Created wheel for flash-attn:
    filename=flash_attn-2.5.7-cp310-cp310-linux_x86_64.whl size=120853563
    sha256=bbe6f77fd0899f8a125a5bdcf734b660c4c88e81c9b51c7ce98ebeba44dc6fa0
  Stored in directory: /root/.cache/pip/wheels/13/96/ed/bcac89c56b606421f99b45b1
6a94db5d0f2b6b4eaf8bac4d01
Successfully built flash-attn
Installing collected packages: ninja, flash-attn
Successfully installed flash-attn-2.5.7 ninja-1.11.1.1

```

```

[ ]: !wget -O image.jpg https://live.staticflickr.com/4071/4245769620_e33561e242_z.
    ↪jpg

```

```

--2024-04-20 15:24:21--
https://live.staticflickr.com/4071/4245769620_e33561e242_z.jpg
Resolving live.staticflickr.com (live.staticflickr.com)... 108.156.94.80,
2600:9000:2351:ac00:0:5a51:64c9:c681, 2600:9000:2351:a00:0:5a51:64c9:c681, ...
Connecting to live.staticflickr.com
(live.staticflickr.com)|108.156.94.80|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [image/jpeg]

```

Saving to: 'image.jpg'

image.jpg [ <=> ] 31.97K --.-KB/s in 0.002s

2024-04-20 15:24:21 (18.0 MB/s) - 'image.jpg' saved [32741]

```
[ ]: from transformers import AutoModelForCausalLM, AutoTokenizer
from PIL import Image
import torch

model_id = "vikhyatk/monodream2"
revision = "2024-03-13"

# Check if CUDA (GPU support) is available and then set the device to GPU or CPU
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

# Load the model and tokenizer
model = AutoModelForCausalLM.from_pretrained(model_id, trust_remote_code=True,
↪revision=revision).to(device)
tokenizer = AutoTokenizer.from_pretrained(model_id, revision=revision)
```

/usr/local/lib/python3.10/dist-packages/huggingface\_hub/utils/\_token.py:88:

UserWarning:

The secret `HF\_TOKEN` does not exist in your Colab secrets.

To authenticate with the Hugging Face Hub, create a token in your settings tab (<https://huggingface.co/settings/tokens>), set it as secret in your Google Colab and restart your session.

You will be able to reuse this secret in all of your notebooks.

Please note that authentication is recommended but still optional to access public models or datasets.

warnings.warn(

config.json: 0%| | 0.00/318 [00:00<?, ?B/s]

configuration\_monodream.py: 0%| | 0.00/3.39k [00:00<?, ?B/s]

monodream.py: 0%| | 0.00/5.41k [00:00<?, ?B/s]

modeling\_phi.py: 0%| | 0.00/49.4k [00:00<?, ?B/s]

vision\_encoder.py: 0%| | 0.00/3.60k [00:00<?, ?B/s]

model.safetensors: 0%| | 0.00/3.72G [00:00<?, ?B/s]

generation\_config.json: 0%| | 0.00/69.0 [00:00<?, ?B/s]

tokenizer\_config.json: 0%| | 0.00/7.34k [00:00<?, ?B/s]

vocab.json: 0%| | 0.00/798k [00:00<?, ?B/s]

merges.txt: 0%| | 0.00/456k [00:00<?, ?B/s]

```
tokenizer.json: 0%|          | 0.00/2.11M [00:00<?, ?B/s]
```

```
added_tokens.json: 0%|          | 0.00/1.08k [00:00<?, ?B/s]
```

```
special_tokens_map.json: 0%|          | 0.00/99.0 [00:00<?, ?B/s]
```

Special tokens have been added in the vocabulary, make sure the associated word embeddings are fine-tuned or trained.

## 1 Image

```
[ ]: image = Image.open('pitch.jpeg') #path of the input image

enc_image = model.encode_image(image).to(device)

# Generate the answer
answer = model.answer_question(enc_image, " why is this edible", tokenizer)

# If the output is a tensor, move it back to CPU for further operations like
↳ print
if isinstance(answer, torch.Tensor):
    answer = answer.cpu()

print(answer)
```

The man in the image is a professional cricketer, and he is taking a swing at the ball with his bat. He is wearing a helmet and is in the middle of a game. The presence of a crowd watching the game indicates that he is playing in a competitive environment, which requires skill, focus, and dedication. Being a professional cricketer, he is responsible for his performance and the performance of his team. The fact that he is taking a swing at the ball suggests that he is trying to score runs for his team, which is why he is playing at such a high level. His presence in the image is a representation of the dedication and commitment required to excel in the sport of cricket.

## 2 GIF

```
[ ]: # from google.colab import files
# uploaded = files.upload()

from PIL import Image
```

```
[ ]: import locale; locale.getpreferredencoding = lambda: "UTF-8"
```

```
[ ]: !wget -O tenor.gif https://media1.tenor.com/m/98GZJjf3CeUAAAAC/bake-pie.gif
```

```
--2024-04-20 15:06:20-- https://media1.tenor.com/m/98GZJjf3CeUAAAAC/bake-
pie.gif
```

```

Resolving media1.tenor.com (media1.tenor.com)... 74.125.203.95, 108.177.125.95,
64.233.187.95, ...
Connecting to media1.tenor.com (media1.tenor.com)|74.125.203.95|:443...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 3408429 (3.2M) [image/gif]
Saving to: 'tenor.gif'

tenor.gif          100%[=====>]    3.25M  --.-KB/s    in 0.03s

2024-04-20 15:06:20 (119 MB/s) - 'tenor.gif' saved [3408429/3408429]

```

```

[ ]: # Load the GIF file
gif_path = 'catch.gif' # Replace with your gif path
gif = Image.open(gif_path)

frame_count = 0
answers = []

while True:
    try:
        # Seek to the next frame
        gif.seek(frame_count)

        # Process every 5th frame
        if frame_count % 5 == 0:
            # Convert the frame to a format compatible with the model
            frame = gif.convert('RGB')

            # Process the frame with the model
            enc_image = model.encode_image(frame).to(device)
            answer = model.answer_question(enc_image, "Give me the description_
of the image and also read the text on the image", tokenizer)
            if isinstance(answer, torch.Tensor):
                answer = answer.cpu()

            # Append the answer
            answers.append(answer)

        frame_count += 1
    except EOFError:
        # Exit the loop when the end of the GIF is reached
        break

# Concatenate and print all answers
all_answers = '\n'.join(answers)

```

```
print(all_answers)
```

The image captures a thrilling moment during a cricket match, where a player is running towards the boundary, attempting to score a run. The scene is set on a lush green field, with several players scattered around the area.

In addition to the players, there are a few cars and a truck parked nearby, possibly belonging to the spectators or the players. The cars are positioned in various spots around the field, while the truck is located closer to the left side of the image. The atmosphere is filled with excitement as the game unfolds. The image captures a thrilling moment during a cricket match, where a player is running towards the boundary, while the opposing team's player is attempting to catch the ball. The action is so intense that a firework can be seen in the background, adding to the excitement.

There are several people in the scene, including the players and possibly some spectators. One of the players is wearing a jacket, which is noticeable in the crowd. The match is taking place on a grass field, and the players are actively engaged in the game.

The image captures an exciting moment during a cricket match, with a player in a blue jersey running towards the ball. The ball is located near the center of the scene, and the player is positioned towards the right side of the image.

There are several other players on the field, some of them closer to the foreground and others further in the background. A few players are also standing near the edges of the image, possibly waiting for their turn to play or observing the ongoing match. The scene is filled with energy and anticipation as the players and the ball are in motion.

The image captures an intense moment during a cricket match, with a player from the opposing team attempting to catch the ball. The player is running towards the ball, while the other player from the same team is trying to catch it. The scene is set on a grassy field, with both players displaying determination and focus on the game.

There are several other players scattered across the field, engaged in various activities related to the match. The image is a snapshot of the action-packed game, showcasing the competitive spirit and athleticism of the players.

The image captures a thrilling moment during a cricket match, where a player is about to run towards the boundary. The player is in the process of taking off from a run-out position, while the other player is waiting for the ball. The scene is filled with anticipation as the outcome of the play hangs in the balance.

There are two other players visible in the scene, one on the left side of the image and another on the right side. A sports ball can be seen in the air, likely the cricket ball that has been hit by the bat. The players are all focused on the game, and the intensity of the moment is palpable.

The image captures a thrilling moment during a cricket match, where a player is in mid-leap, attempting to catch the ball. The player is crouched down, extending his leg to reach the ball, while the rest of the team watches the play unfold.

The scene is set on a lush green field, with the players and the ball visible in the image. The atmosphere is intense as the players and the audience are captivated by the ongoing match.

The image captures a thrilling moment during a cricket match, where a player is in the process of throwing up his hands in frustration. The player is standing on the field, and the scene is filled with excitement as the game unfolds.

The grassy field is well-maintained, and there are multiple people present, including the player who threw up his hands and other teammates and opponents. Some of the players are positioned closer to the foreground, while others are scattered across the field.

In addition to the players, there are a few bottles placed around the field, likely containing water or sports drinks for the athletes to stay hydrated during the match.

The image captures a thrilling moment during a cricket match, where a player is leaping into the air to catch a ball. The player is in the process of reaching out to catch the ball, which is in mid-air.

The scene is set on a field with a few other players and a sports ball. There is a bench located near the edge of the field, and a bottle can be seen placed on it. The image is filled with excitement and action, showcasing the intensity of the game.

The image captures a thrilling moment during a cricket match, where a player is in mid-air, leaping to catch a ball. The player is wearing a blue and yellow uniform and is skillfully leaping over another player to make the catch.

The scene takes place on a grassy field, with a crowd of spectators watching the game intently. There are multiple people in the background, some of them closer to the action, while others are further away. A bottle can be seen on the ground near the edge of the field, possibly belonging to one of the players or spectators.

The image captures a thrilling moment during a cricket match, where a player is in mid-air, attempting to catch a ball. The player is jumping off the ground with his hands outstretched, showcasing his athleticism and skill.

The scene takes place on a field with a large advertisement banner in the background, likely promoting a sports brand or event. The advertisement is visible in the upper part of the image, emphasizing the connection between the exciting game and the brand's marketing efforts.

The image captures an exciting moment during a cricket match, where a player is in mid-air, jumping to catch a ball. The player is wearing a black and yellow uniform and is skillfully leaping off the ground to make the catch.

The scene takes place on a grassy field, with a fence in the background. There are several other people present in the image, likely teammates or opponents, watching the action unfold. Some of them are standing closer to the fence, while others are positioned further back. The atmosphere is filled with anticipation and excitement as the player attempts to catch the ball and potentially score a run for their team.

The image captures an intense moment during a cricket match, where a player is diving to catch a ball. The player is in mid-air, displaying their athleticism and skill. The scene is set on a grass field, with a fence in the background.

There are several people present in the scene, including teammates and opponents, all focused on the ongoing match. A sports ball can be seen in the air, close to the diving player. Additionally, there is a bottle placed on the ground, likely belonging to one of the players or spectators.

The image captures an intense moment during a cricket match, where a player is falling to the ground. The player is in the process of getting up, while another player is laying on the ground nearby. The scene takes place on a grassy field, with a fence in the background.

There are several people in the scene, including a player who is laying on the ground, another player who is getting up, and a few others scattered around the field. A sports ball, likely a cricket ball, can be seen in the air, indicating that the match is still ongoing.

The image captures a thrilling moment during a cricket match, where a player is falling onto the ground after a spectacular catch. The player is in the process of sliding into the catch, showcasing their athleticism and skill.

The scene takes place on a grassy field, with a crowd of people watching the game intently. There are at least ten individuals visible in the image, some of them standing closer to the action, while others are further away. The atmosphere is filled with excitement and anticipation as the player's catch unfolds.

The image captures a thrilling moment during a cricket match, where a player is sliding into a catch to avoid being caught by the opposing team. The player is in the process of getting up, while the other players on the field are holding their positions, waiting for the next move.

The scene is set on a grass field, with a large crowd watching the match intently. There are several players visible in the image, with some standing closer to the action and others further back in the field. The atmosphere is filled with excitement and anticipation as the game unfolds.

The image captures a thrilling moment during a cricket match, where a player is diving to catch a ball. The player is lying on the ground, wearing a blue shirt and white pants, as he attempts to catch the ball.

The scene takes place on a grassy field, with a large crowd of spectators watching the game intently. There are multiple people in the background, some of



them standing and others sitting, all focused on the action unfolding before them. The atmosphere is electric, as the players and the audience alike are captivated by the intense competition.

The image captures a thrilling moment during a cricket match, where a player is sliding into a catch to avoid being out. The player is wearing a 50 jersey, which indicates their team's jersey number. The scene takes place on a grass field, with the player's teammates and opponents watching the intense play unfold.

There are several people in the image, including the player sliding into the catch, other players on the field, and spectators. A sports ball, likely a cricket ball, is visible in the scene, adding to the excitement of the moment. The image is a visual representation of the action and emotion that occurs during a live cricket match.

The image captures a thrilling moment during a cricket match, where a player is lying on the ground, possibly after a fall. The player is wearing a number 50 jersey, and the scene is set on a grass field.

There are several other people in the scene, including teammates and opponents, who are watching the event unfold. Some of them are standing closer to the player who fell, while others are positioned further away. The atmosphere is filled with excitement and anticipation as the match continues.

The image captures a thrilling moment during a cricket match, where a player is lying on the ground after a fall. The player is wearing a blue hat and is positioned near the center of the scene. There are two other players visible in the image, one on the left side and another on the right side of the scene.

The match is taking place on a grass field, and there are several potted plants scattered around the area, adding a touch of greenery to the environment. Additionally, there are two bottles placed on the ground, likely belonging to the players or spectators.

The image captures a thrilling moment during a cricket match, where a player is lying on the ground, possibly after a fall. The player is positioned near the center of the scene, while another player stands nearby, watching the event unfold.

The match is taking place on a grass field, and there are multiple benches scattered around the area, likely for the players to rest between innings. A bottle can be seen on the ground, possibly belonging to one of the players or spectators. The atmosphere is intense as the players and the audience eagerly await the outcome of this dramatic play.