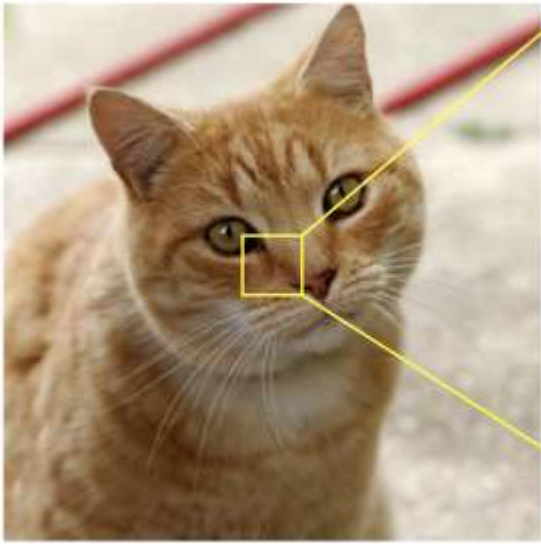# Y-Detective Image and Language AI

Team leader: Xinyang (Andy) Chen

Other members: Bingyu W, Enrique Z, Zhiyu P. W, Weihao J, Tianrui W

We utilize AI to expand human limits.

Contents

- **Inspirations, ideas, and goals**
- Project details and technology utilization
  - Object positioning and cropping
  - Image selection by user descriptions
  - Image class management
  - Other Clarifai tools
- Potential applications and business values
- Future improvements

- Combining computer vision and natural language processing, we are passionate about building smart apps to increase productivity.

- Our goal is to assist human workers in image detection and selection tasks, enabling higher efficiency and quality, thereby reducing the amount of human resources required.

- Eventually, we expect to deliver social benefits and market values.

Contents

- Inspirations, ideas, and goals

- **Project details and technology utilization**
    - **Object positioning and cropping**
    - **Image selection by user descriptions**
    - **Image class management**
    - **Other Clarifai tools**

- Potential applications and business values

- Future improvements

Object detection, positioning, and crop generation

Technology: GPT-4 Vision

What users provide:

- An image
- Descriptions of one or more objects in the image
- Rough number of all objects to focus on

What we return:

- The original image with detected objects labeled in red rectangles
- Cropped images saved as new files
- AI explanation about where the objects are

Crop *"Red bus route"* from the image

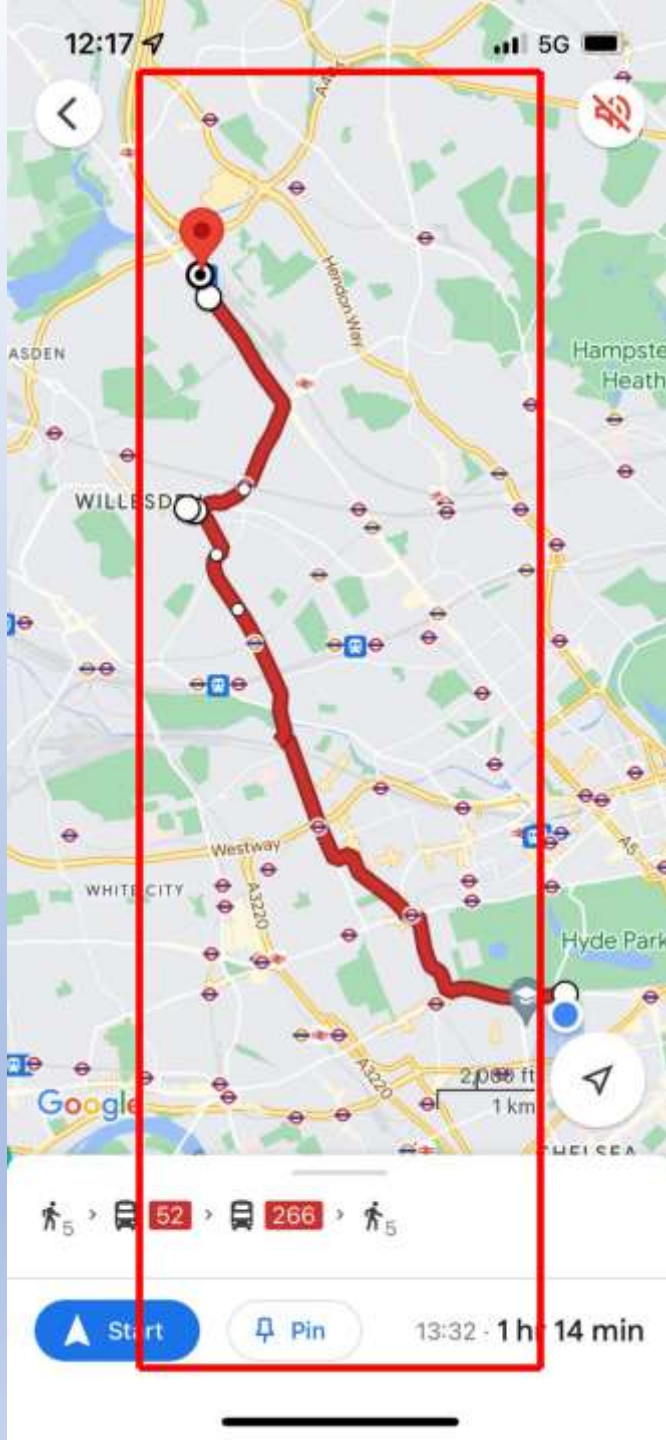# 1. Automatic highlighting and cropping in seconds

2. Explanation provided

red bus route

The red bus route in the image starts approximately at the top and ends near the bottom of the image. It spans nearly the entire height of the image, but it is not perfectly vertical, so the width coverage is not consistent throughout. However, I will provide an estimation that encompasses the entire route:

- Width: 20% to 80% (to cover the entire width of the route at its widest point)
- Height: 5% to 95% (to cover the entire length of the route from top to bottom)

Please note that these percentages are approximations and the actual values may vary slightly.

3. Supporting multiple object cropping in one API call

Technology: GPT-4 Vision

What users provide:

- An album containing images of a certain class
- What common features these images have
- Descriptions of the image(s) to select from the album

What we return:

- A snapshot of all images in the album align as a grid with index labels
- All images matching the user's descriptions
- AI explanation about why these images are selected

Y-Detective AI image operation and detection tools | More applications

Find a set of useful objects from an image

Upload an image to start: [Choose File] No file chosen
Describe all objects to emphasize. If belonging to the same category, describe the category: e.g., human faces, id cards with photo and name, etc.

Roughly how many objects to find in total? ○ Single (1); ○ 2-4; ○ 5-10; ○ > 10; [Find objects]

Find an image matching your descriptions most closely

Enter or select the class of image to find through:
[ ] (Not selected, 0 images) ∨ View and edit image classes
Describe the image you are looking for:

[Find image]

Result overview [clear history]

static/imggroups/maps/IMG_20230914_182014.jpg

Summary and explanations

Find *"Arabic grand mosque"* from the selected album

1. One API call scans through all images
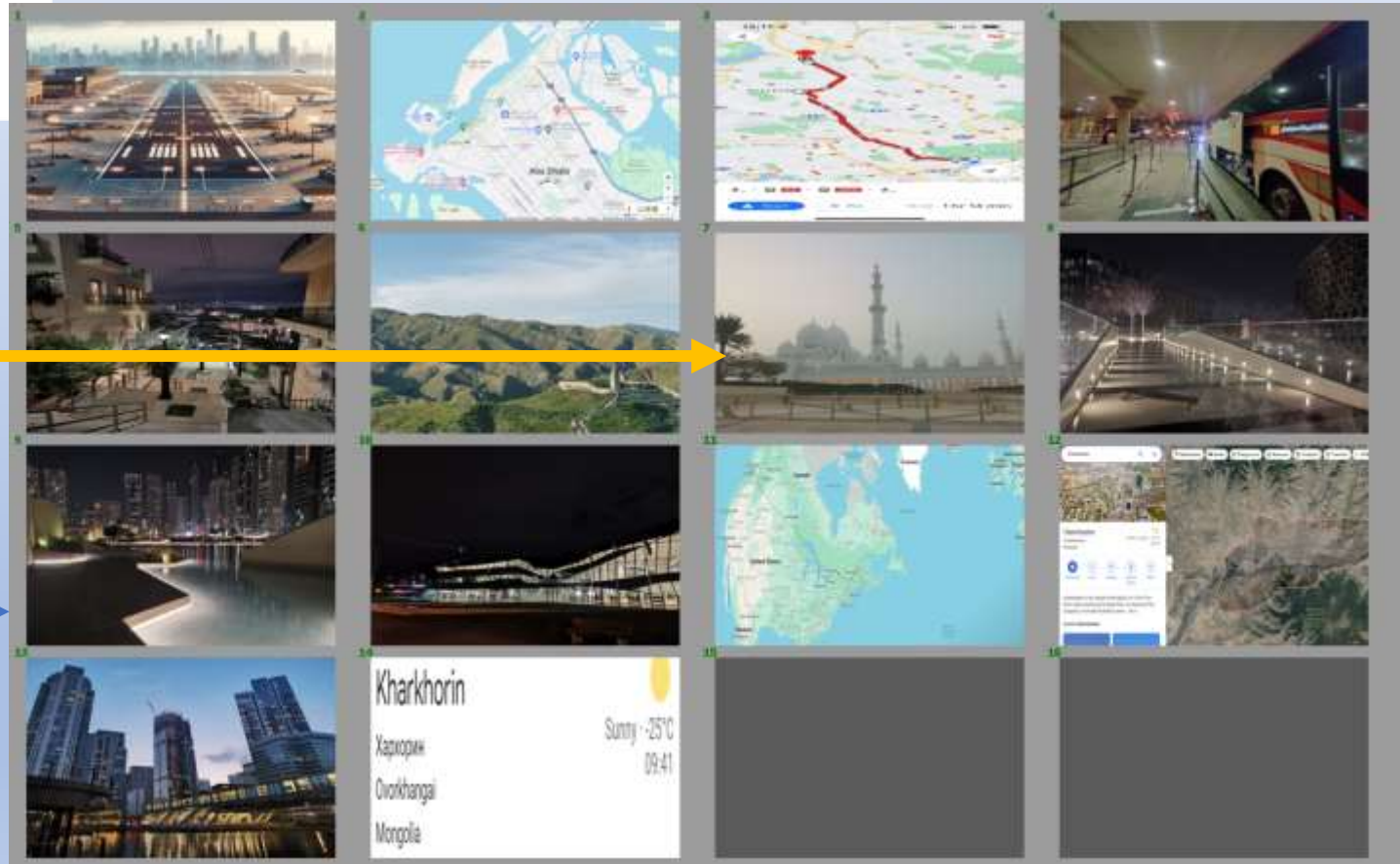
Combine a list of image files to a single snapshot with labels
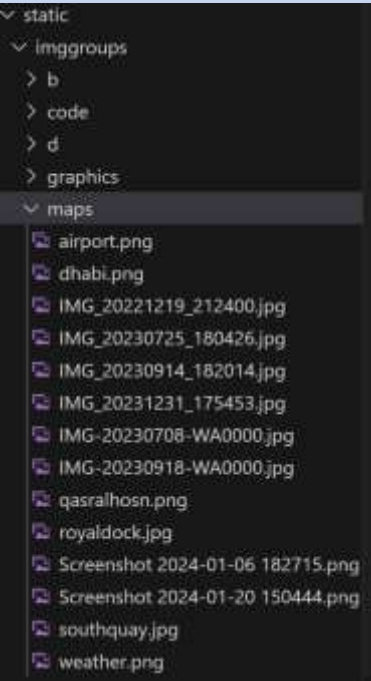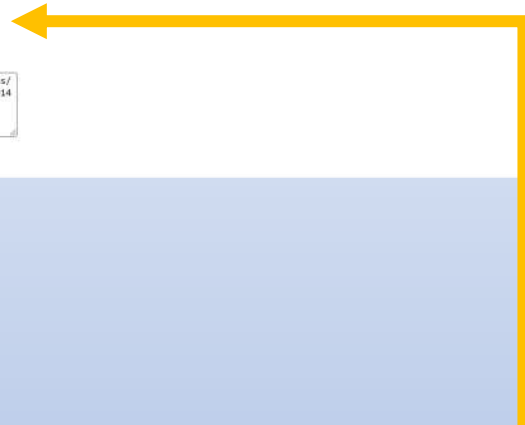
Summary and explanations

arabic grand mosque. All small pictures have common features: maps, urban, traveling, geography

Number 7

The image indexed as Number 7 shows a grand mosque with distinctive Arabic architecture, featuring minarets and a large dome, which matches the description of an "Arabic grand mosque."

2. Clear numbering and explanation provided
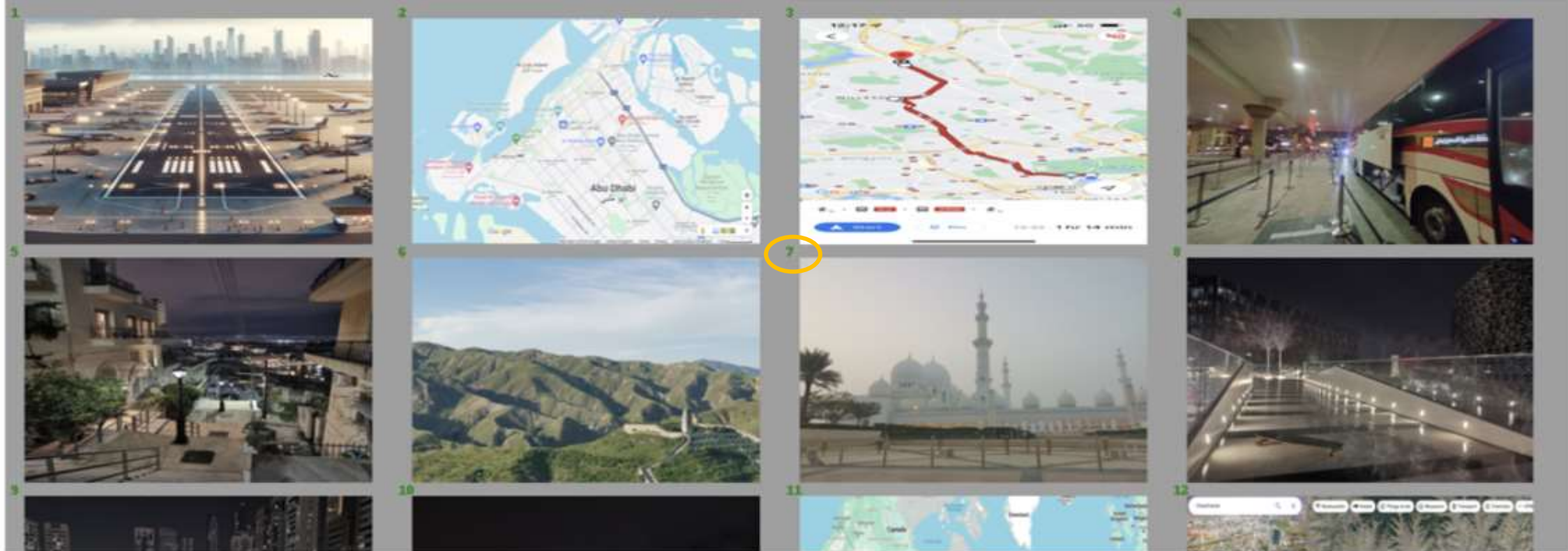
3. Supporting multiple image selection

More than 1 image might match the user's descriptions
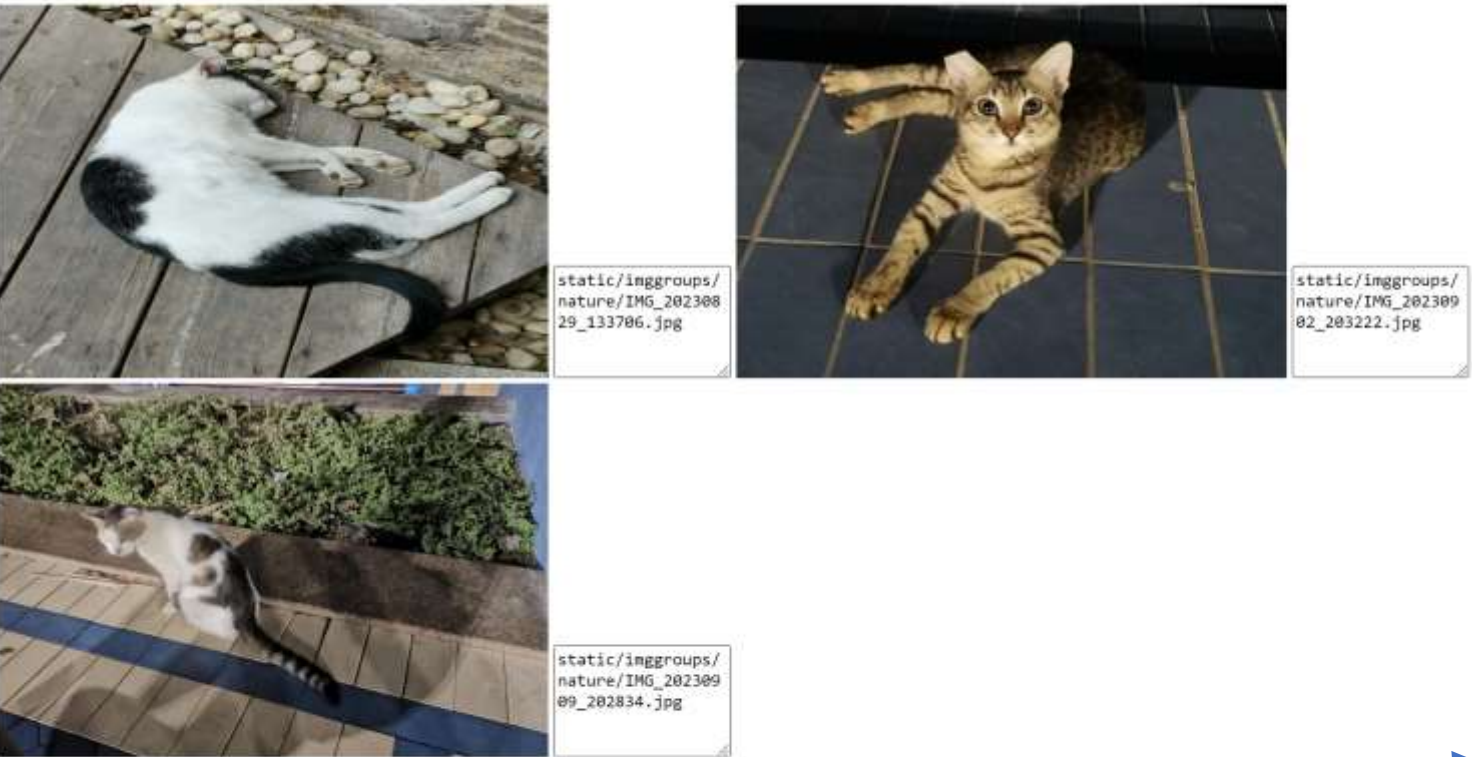
cat resting on the ground. All small pictures have common features: nature, animals, living creatures, exploration

Number 2, Number 3, Number 4

Explanation of the selection:
- Number 2 shows a cat resting on the ground, which matches the description.
- Number 3 also shows a cat lying on the ground, looking up at the camera.
- Number 4 depicts a cat that appears to be resting or walking on a ledge, which is close enough to the description of a cat resting on the ground.

Manage albums: image classes

Manage image classes | Home: AI image tools

Create a new image class [Create class +]

new class name:	
class descriptions:	

Edit image class | [See all image classes >>]

Select/Enter an images class: (Not selected, 0 images) ▼ [View class]

Actions	original	new
rename class	None	
change descriptions		

[Remove class -] [Delete all images -] [Confirm changes]

All image classes [Clear all classes -]

class name	descriptions
code	programing, science, technology, website
maps	maps, urban, traveling, geography
graphics	is a square filled with different colors
nature	nature, animals, living creatures, exploration
b	example to be added
d	

```
∨ static
  ∨ imggroups
    > b
    > code
    > d
    > graphics
    > maps
    > nature
```

1. Each class of images is stored in a sub-directory

2. Class names and descriptions are recorded in a database, easy creation/removal/updates.

Other tools using Clarifai

1. Image generation from user prompt

"City night with skyscrapers and railway bridge over water"

Technology: DALL-E 3

image discussion, image generation, and text-to-speech | Home: AI image tools

If you want to ask questions about an image, upload here (optional)

Choose File | No file chosen -- from "static/imgorig" directory

Describe your question/prompt:

Expect token limit (optional, default 250):
Name the result image/audio (optional): | Generate image | Ask question | Read the text aloud

Question/Requirement:

What should I wear for outside today?

AI answer to the question
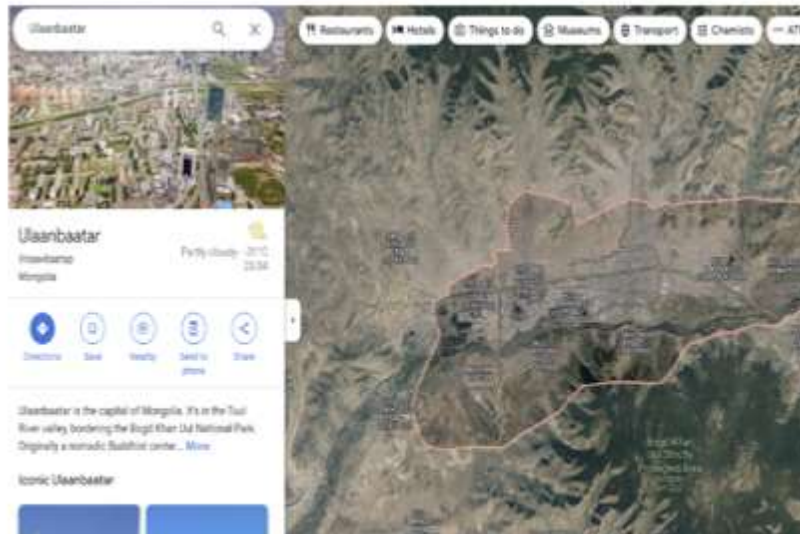
Based on the weather information provided for Ulaanbaatar, Mongolia, showing a temperature of -31°C and partly cloudy conditions, you should dress very warmly to protect yourself from the extreme cold. Here are some clothing recommendations:

1. Thermal underwear: Start with a base layer of thermal underwear, which will help retain body heat.
2. Insulating layers: Add one or more insulating layers, such as fleece or wool sweaters, to provide additional warmth.
3. Outer layer: Wear a heavy, insulated coat that is windproof and waterproof to protect against the

2. Ask questions about an image

Technology:

GPT-4 Vision
GPT-4 Turbo

3. Text-to-speech generation

image discussion, image generation, and text-to-speech | Hom

If you want to ask questions about an image, upload here (optional)

Choose File | No file chosen   -- from "static/imgorig" directory

Describe your question/prompt:

Expect token limit (optional, default 250):
Name the result image/audio (optional): | Generate image | Ask question | Read the text aloud |

Question/Requirement:

Tidal energy is a form of power produced by the natural rise and fall of tides caused by the gravitational interaction between Earth, the sun, and the moon. Tidal currents with sufficient energy for harvesting occur when water passes through a constriction, causing the water to move faster.

Result speech audio

▶ 0:00 / 0:18 ────── 🔊 ⋮

Technology:
Open AI text-to-speech

Contents

- Inspirations, ideas, and goals

- Project details and technology utilization
    - Object positioning and cropping
    - Image selection by user descriptions
    - Image class management
    - Other Clarifai tools

- **Potential applications and business values**

- Future improvements

Media production/editing

- Automated, scalable, and parallelized photo cropping and searching, saving a lot of human work

Customer assistance

- Communicate with customers in both language and image, better understanding issues and providing solutions

Scientific research (astronomy, geology, biology…)

- Detect patterns from telescope photos, satellite maps, medical images, etc., or find the images matching certain criteria--tasks sometimes hard for scientists

Crime detection/investigation

- Filter photos of criminal suspects from a wide range of information sources; even generate images of unseen criminals/scenes based on witness reports
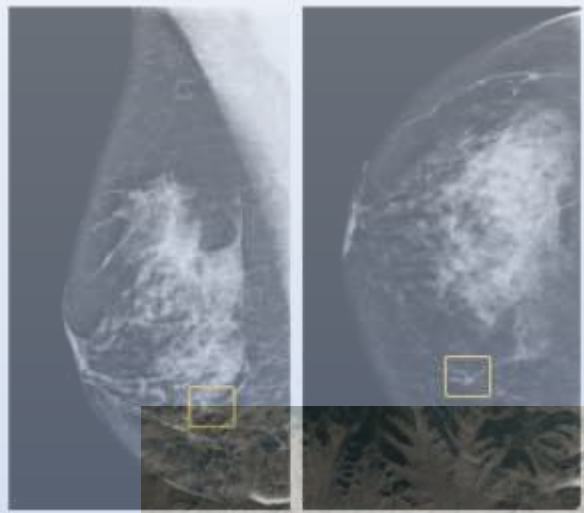
Contents

- Inspirations, ideas, and goals
- Project details and technology utilization
  - Object positioning and cropping
  - Image selection by user descriptions
  - Image class management
  - Other Clarifai tools
- Potential applications and business values
- **Future improvements**

- Improve the accuracy of object positioning: being able to crop all the required items nicely in the center and covering all corners
- Allow multi-level and batch image selection: e.g., dividing a large set of images to 20 batches, for each one find the best matches, and then compare the results
- Optimize file system management: enabling users to easily upload and save images
- Reduce the waiting time for users to receive results

Thank you for learning about our project!

Image & Language AI assistance program

**Y-Detective | **Utilize AI to expand human limits