Flask Application Documentation

Overview

This Flask application handles the upload and processing of ZIP files containing images and their corresponding labels. The application uploads the ZIP files to a MongoDB database, extracts the contents, distributes the files into training, testing, and validation sets, and then uploads them to an AWS S3 bucket.

Dependencies

- Flask
- pymongo
- boto3
- dotenv
- zipfile
- bson
- io
- random

Routes

1. /upload-zip [POST]

Endpoint to upload a ZIP file.

Request:

- image-zip file

Response:

- 200 OK if the file is successfully uploaded to MongoDB.
- 400 Bad Request if there is an issue with the file.

Function:

- Checks for the presence of a file in the request.
- Validates if the file is a ZIP file.
- Reads the file as binary data and uploads it to MongoDB.

2. /process-zip/<file_id> [GET]

Endpoint to process the uploaded ZIP file.

Parameters:

- file_id: The ID of the file in MongoDB.

Response:

- 200 OK if processing and uploading to S3 is successful.
- 404 Not Found if the file is not found in MongoDB.

Function:

- Retrieves the ZIP file from MongoDB.
- Extracts the contents of the ZIP file.
- Creates directories in S3.
- Distributes the files into training, testing, and validation sets.
- Uploads the files to the corresponding directories in S3.

Helper Functions

extract_zip(zip_data)

Extracts the contents of a ZIP file to /tmp/hpc_data_joon.

Parameters:

- zip_data: Binary data of the ZIP file.

Returns:

- A generator yielding the names of extracted files.

upload_to_s3(file_path, s3_path)

Uploads a file to the specified path in the S3 bucket.

Parameters:

- file_path: Local file path.
- s3_path: S3 destination path.

create_s3_directories()

Creates necessary directories in the S3 bucket.

Directories Created:

- test/images/
- test/labels/
- training/images/
- training/labels/

- validation/images/
- validation/labels/

Example Usage

1. Upload ZIP File:
   curl -F 'image-zip=@path/to/your/file.zip' http://localhost:5000/upload-zip

2. Process ZIP File:
   curl http://localhost:5000/process-zip/<file_id>

Replace <file_id> with the actual ID of the uploaded ZIP file from MongoDB.