Image Recognition with IBM Cloud Visual Recognition

921821104037: S. Sarathrajan

Phase 1

Overview

Phase 1 document for the project **Image Recognition with IBM Cloud Visual Recognition** from IBM on NAN MUDHALVAN scheme

Project Title

Image recognition

Project Description

Develop an image recognition system using IBM Cloud Visual Recognition. Share your passion for photography by uploading images and watch as the system accurately classifies and describes their contents. Craft engaging visual stories with the help of AI-generated captions. Connect with your audience through captivating visuals and compelling narratives!

Project Definition

The project involves creating an image recognition system using IBM Cloud Visual Recognition. The goal is to develop a platform where users can upload images, and the system accurately classifies and describes the image contents. This will enable users to craft engaging visual stories with the help of AI-generated captions, enhancing their connection with the audience through captivating visuals and compelling narratives.

Design Thinking

1.Image recognition setup : Sign up for an IBM Cloud Account: If you don't already have one, go to the IBM Cloud website (https://cloud.ibm.com/) and sign up for an account.Log In: After creating an account, log in to your IBM Cloud account.Create an IBM Watson Visual Recognition Service: Inside your IBM Cloud dashboard, navigate to the catalog or services section. Look for "Watson" services and find "Visual Recognition.

" Click on it.Configure Your Service: Follow the on-screen instructions to configure your Visual Recognition service. You may need to specify the service name, plan (e.g., Lite, Standard), and other details.Create Service Credentials: Once your service is created, you'll need to create service credentials to obtain the API keys. Go to the "Service credentials" section of your Visual Recognition service and click on "New credential. Obtain API Keys:

The service credentials you created will contain the API keys. Click on the newly created service credential to view the details, including the API key and other necessary information.Please note that the specific steps and the layout of the IBM Cloud dashboard may change over time, so it's a good idea to refer to the official IBM Cloud documentation for the most up-to-date instructions on setting up the Visual Recognition service and obtaining API keys.

2.User interface: Designing a user-friendly interface for image upload and AI-generated caption viewing involves a combination of aesthetics and functionality. Here's a simple outline for such an interface:

Image Preview:After successful upload, display a thumbnail or preview of the uploaded image.

Allow users to crop or resize the image if necessary.

Caption Generation:once the image is uploaded, display a loading spinner or progress bar to indicate AI processing.

When the AI generates the caption, display it prominently near or below the image.Include an option to generate a new caption in case the user is not satisfied with the first one.

Caption Sharing:Offer options to share the generated caption on social media or copy it to the clipboard.Provide a “Download Caption” button to save it as a text file.

User Profiles (Optional):Allow users to create profiles to save uploaded images and captions.Include a “My Uploads” section for easy access to previously generated captions.

Responsiveness:Ensure the interface is responsive and works well on both desktop and mobile devices

Error Handling:Provide clear error messages and guidance if something goes wrong during the upload or caption generation process.

Privacy and Data Handling:Clearly communicate your privacy policy and data handling practices.Allow users to delete their uploaded images and captions if they wish.

User Feedback:Include a feedback mechanism for users to report issues or provide suggestions.

Aesthetics:Use a visually appealing design with a pleasing color scheme and typography.Ensure that the interface elements are well-organized and easy to navigate.

Accessibility:Make sure the interface is accessible to users with disabilities, including alt text for images and keyboard navigation.

Documentation:Provide clear instructions on how to use the service, including FAQs if necessary.

3.image classification:

API Credentials:After creating the service, you'll need to obtain API credentials (an API key) to authenticate your requests.Collect and Label Images:Prepare a dataset of images for your classification task.Label these images to create classes/categories that you want the model to recognize.Train Your Custom Model (Optional):If you have a specific use case, you can train a custom model using your labeled data. This step is optional, as IBM provides a general pre-trained model as well.

Use the API for Classification:You can use the API to classify images. There are different methods to do this depending on your preference and programming language. Here's a simplified example using Python:import requests

```
Api_key = 'YOUR_API_KEY'
url = 'https://api.us-south.visual-recognition.watson.cloud.ibm.com/…/v3/classify?version=2018-03-19'
# Replace 'YOUR_API_KEY' and the URL with your API key and endpoint.
Headers = {
  'apikey': api_key,
}
# Prepare the image for classification.
With open('image.jpg', 'rb') as image_file:
  Files = {'images_file': image_file}
# Send the request for classification.
Response = requests.post(url, headers=headers, files=files)
# Parse and interpret the classification results from the response.

Results = response.json()
```

Interpret Results:The results will contain information about the classes/categories detected in the image.

Handle Errors and Edge Cases:Be prepared to handle cases where the API might not be able to classify an image accurately.

Deploy and Integrate:Integrate this API into your application or system as needed.

4.AI-Generated caption : Integrating natural language generation (NLG) with image recognition can be a powerful way to automatically generate captions for recognized images. Here's a high-level overview of how you can achieve this:

Image Recognition: Use a computer vision model or API, such as a convolutional neural network (CNN) or a pre-trained model like Google Vision API or Microsoft Azure Computer Vision, to recognize objects and scenes in the images. This step will provide you with a list of detected objects and their positions within the image.

Data Preparation: Extract the relevant information from the image recognition results, such as the detected objects, their locations, and any other pertinent data like colors or attributes.

Natural Language Generation (NLG): Implement an NLG model, such as a recurrent neural network (RNN), transformer-based model, or GPT-3 (like this one), to generate natural language captions based on the detected objects and additional image data. You can fine-tune a pre-trained language model on a dataset of image-caption pairs to improve caption quality.Template-Based or Neural Approach: You can take one of two approaches to NLG:

Template-Based: Create a set of predefined caption templates and fill in the detected objects and attributes from the image recognition results. This approach is more rule-based but can be effective.

Neural NLG: Train a neural network to generate captions from scratch based on the input from step 2. This approach provides more flexibility but requires a larger amount of training data and computational resources.

Combine and Refine: Combine the generated caption from the NLG model with any additional context or information. You may also apply post-processing to ensure grammatical correctness and coherence.

Testing and Optimization: Thoroughly test the system to ensure that it produces accurate and meaningful captions. Fine-tune both the image recognition and NLG components as needed to improve the quality of the captions.

Integration: Integrate this system into your application or platform, ensuring that it can accept images as input and provide generated captions as output.

5.User engagement : Gallery View: Provide a gallery view where users can see a grid of their AI-enhanced images. Each image should have a thumbnail and a title or description.

Upload and Import: Allow users to upload images from their device or import them from cloud storage services like Google Drive or Dropbox.

AI Enhancement Options: Provide a variety of AI enhancement options such as image filters, style transfers, color adjustments, and object recognition. Users should be able to apply these enhancements to their images easily.

Save and Organize: Enable users to save their enhanced images and organize them into albums or folders. This helps users keep their images well-organized and easily accessible.

Sharing Options: Implement sharing options, including social media integration (e.g., Facebook, Instagram, Twitter), email sharing, and generating shareable links. Users should have the flexibility to choose their preferred sharing method.

Privacy Settings: Allow users to set privacy preferences for their images. They should be able to choose whether their images are public, private, or shared with specific individuals or groups.

Collaboration: If applicable, enable collaborative editing, where multiple users can work on the same image and see real-time changes.

Comment and Like: Include features for users to leave comments and like or favorite images within the app to encourage social interaction.

Search and Filter: Implement search and filtering options to help users quickly find specific images within their gallery, based on keywords, dates, or tags.

Image Information: Display image metadata such as date created, location (if available), and the AI enhancement techniques applied.

Notifications: Send notifications when someone likes or comments on a user's shared images or when they receive shared images from others.

Offline Access: Allow users to access their saved images even when they are offline by caching them locally on their devices.

Cross-Platform Compatibility: Ensure that the app works seamlessly across various platforms, including smartphones, tablets, and desktops.

Backup and Sync: Offer an option for users to back up their image data and sync it across devices to prevent data loss.