Scene Pictures Classification and Captioning

Riya Patel, 001198749
BSc Computer Science (Hons)
University of Greenwich

Contact Information:
Computing and Mathematical Science
University of Greenwich
SE10 9LS, London, UK

Email: riya.mpatel77@gmail.com

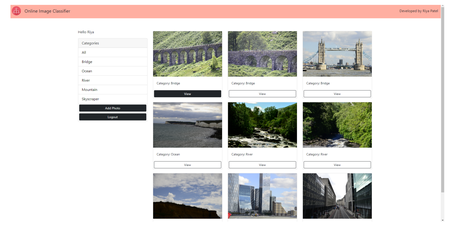


Figure 1: Album page

Abstract

A classification and captioning application has been developed to categorize pictures taken by photographers. Due to the large volume of unclassified and undefined photos, it is impractical for users to label each one individually. To address this, an app has been created that allows users to upload their pictures. The app has a user interface that enables users to join and automatically sorts the uploaded images into different classes. Additionally, the application provides captions for each picture. The classification algorithm was trained on 10 classes and demonstrated good results. For captioning, different models were analysed to label images. By using different machine learning techniques with a user interface, a model has been built that can assist with repetitive tasks.

Introduction

Image classification, has gained significant attention over the past decade. This technology is used in many applications, from autonomous vehicles that require rapid and accurate image recognition for safe navigation to social media platforms like Facebook and Google Photos that use these techniques to enhance user experience.

Over 1.72 trillion photos are being taken annually, the demand for efficient and accurate image classification algorithms has become critical to group these images. The ability to effectively manage and use this big volumes of images have implications across various sectors. In the medical field, precise image analysis can aid in early disease detection, while in the retail sector, it can help in inventory classification, saving time and increasing revenue. The ultimate goal is to develop a sophisticated system that can categorize and caption photographs across diverse categories by using machine learning. This would help the way we organize and locate specific images.

Project Idea and Objectives

The objective of this project is to provide photographers with a platform where they can upload their images, and our system will classify them into categories. To achieve this, the first step is to develop a user-friendly interface that allows users to interact with the application. First a main page is needed to increase the website visibility and accessibility. Furthermore, users should be able to create accounts and log in securely. This will enable them to access their accounts and see their images while maintaining their data privacy. By creating accounts, users can manage and organize own image collections. Once registered, users should be able to upload their images to the application. These images will be associated with their respective accounts. Another requirement is to let users to select specific categories that they want to view. This feature will enable users to filter and organize their image collections based on the categories on their album.

To achieve the aims, several objectives must be met..

1. This study investigates various image classification algorithms, considering factors such as accuracy, performance, scalability, and computational resource requirements. The objective is to identify an algorithm that can efficiently and accurately categorize images based on their visual content and characteristics.
2. The image classification model will be trained on a collection of chosen categories. This selection is needed for the model to learn and recognize patterns and scenes accurately. By showing the model to a wide range of images that photographers take, it will return better results.
3. A user-friendly web interface will be designed and implemented to provide photographers a better experience. The interface will allow users to create accounts, securely log in, upload images, and filter their collections based on categories. This will enable photographers to easily organize and access their images, saving time and effort in managing their collections.
4. The image classification model will be trained with different photographs to get a high accuracy. With the results an analysis will be made on how the model performs with different scenes. Moreover, potential challenges related to image classification, such as overfitting will be investigated and addressed. By solving these issues, the systems robustness and reliability will be improved.
5. Finally, the project explores different implementations of image captioning models including pre-trained models to describe images automatically.

Product

Web application

A web application was developed using Django to allow users to upload, manage, and interact with their photo collections. The application offers several key features to enhance user experience and functionality. Firstly, it provides a secure login system that enables users to access their personal accounts and view their uploaded images. This ensures the privacy and safety of their photo collections. Secondly, the application has tools to upload and manage pictures. Users can add new photos with a class and description. Lastly, the application includes a sorting feature that categorizes the pictures based on the classes.

Image classification

An image classification algorithm is implemented to automatically categorize uploaded images. When the user does not specify a class, the algorithm analyzes the image and assigns one category.
The system has been trained on the Places205 dataset. However, the model was trained with 10 scene classes: sky, ocean, mountain, river, bridge, harbor, rainforest, skyscraper, coral reef, and castle. Each class contains 2,000 images.
During the development process, four different CNN were built and evaluated. After comparing their performance, the best model was selected for implementation in the final system.

Image Captioning

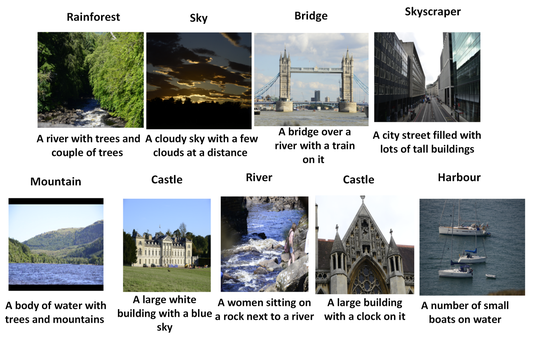
In addition to image classification, the application also has an image captioning functionality. The image captioning uses a pre-trained model based on the transformer architecture. When a user uploads an image without a description, the model analyzes the content of the image and generates a caption that describes the elements in the photograph.

Results

The results indicate that the majority of the classes are accurately classified, and the captioning is correct in most instances. However, there are some inaccuracies that require more research. These can be because of the lack of sufficient data for training the model, particularly when it comes to scene recognition. The limited availability of diverse and representative training data hinders the model's ability to generalize and accurately classify and caption images across a wide range of scenarios. To improve the performance and robustness of the model, it is crucial to expand the training dataset, incorporating a broader variety of scenes and objects. This will enable the model to learn more comprehensive features and patterns, ultimately leading to enhanced accuracy in both classification and captioning tasks.

	Accuracy	Precision
CNN1	0.60	0.61
CNN2	0.63	0.64
CNN3	0.60	0.61
CNN4	0.68	0.68

Table 1: Places205 dataset results with 2000 pictures for each class



Conclusions

This project has showed many challenges and opportunities in the field of image classification and captioning, despite the extensive research already conducted.

- The practical applicability of these techniques is limited, and image captioning, in particular, is a feature that is not yet widely adopted in many applications.
- In conclusion, our work has demonstrated the potential for further advancements in image classification and captioning, emphasizing the need for applicability and development in these areas to get more products into real-world applications.

Limitations Further Work

- When attempting to train the model with more than 25,000 images, the runtime consistently crashed, requiring the model to be restarted.
- To determine this limit, we incrementally increased the number of training images and observed that the model crashed every time the image count surpassed 25,000.
- This limitation may restrict the applicability of our developed model to a broader range of scene types and photographic collections.
- To address this limitation and expand the scope of our work, future research should consider incorporating additional classes from the Places205 dataset.
- For image captioning, to develop a better captioning model, scene recognition can be used instead of object detection.
- To improve user experience, implementing sorting options and allowing users to edit descriptions. These features are easy to use and would provide better user experience.
- Furthermore, future work should explore the potential of integrating other datasets or combining multiple datasets to further diversify the training data.

  Hence, by expanding the range of scene categories, and incorporating additional datasets, we can develop more comprehensive and robust models that better serve the needs of photographers in organizing and categorizing their collections.