January 2022

Determine Results for the People-Related Image Queries Through Object Detection

Jing Xiong

DETERMINE RESULTS FOR PEOPLE-RELATED IMAGE QUERIES THROUGH OBJECT DETECTION

Introduction

The present disclosure relates to computer-implemented systems and methods for generating image search results relating to objects within a provided image. More particularly, embodiments of the present disclosure allow for returning search results relating to objects detected within images that have been deemed as private or containing sensitive information (e.g. an image from a family vacation to Disney World with a minor present). When provided an image to search, the system looks for any potentially sensitive information such as, for example, a person's face and filters out the information. The system then analyzes the rest of the image, absent the sensitive information, to determine potential search results to return the user.

Traditionally, visual search applications do not show results at all when receiving a query with images that contain sensitive information such as, for example, a person's face because showing similar images results that contain similar faces may generate results that cause the user to feel uncomfortable or generate results that the person who's said face belongs to would not like to be a search result for any user in the world. For example, if a user searched their own photo or a relative's, and retrieved criminals' photos as similar image results or a sensitive photo of the person the user searched for (i.e. a mugshot, picture from an old relationship, etc.), it would be an unpleasant or uncomfortable situation to for the user. The systems and methods described in the present application provide the added effect and benefit of preventing that while still returning helpful or desired search results.

Summary

Computer-implemented systems and methods for generating image search results in accordance with the disclosed technology can generate image search results from query images containing sensitive information. In some instances, a user may provide a query image containing sensitive information such as, for example, a minor's face. Utilizing the face detected within the image to generate image search results may unintentionally cause privacy concerns for the minor and potentially cause the user discomfort. Therefore, using the disclosed techniques can eliminate the issue by filtering out sensitive information detected within a given query image.

More particularly, a query image can be provided to a first layer of a machine learned model that is capable of detecting, for example, if an image contains people sensitive information such as, for example, a person's face. The results of the first layer can then be passed to a second layer of the machine learned model to determine if the image contains information that is not people sensitive such as, for example, a famous building, an article of clothing, an instrument, etc. From there, search results based on the non-people sensitive information can be generated. In some implementations, the image quality, people sensitive information, and non-people sensitive detection confidence can be checked before generating search results. If the detection confidence and/or image quality are above a pre-determined threshold, search results will be generated. If the parameters are below the pre-determined threshold, the user will be returned an error message to reduce and prevent the possibility of people sensitive information being used to generate search results (i.e. not returning results if the model does not return a high level of confidence in blocking out sensitive information).

In another embodiment, the user query image can be input to a first layer of a machine learned model that is capable of detecting, for example, a person's face, and returning a region of

the image that should not be used to search for objects of interest or intension of search results. Once the result of the first layer of the machine learned model is returned, the query image can be passed to a second layer of the machine learned model which utilizes both the query image and the results from the first layer of the machine learned model to determine search results from the query image. These search results would not contain any results relating to the potentially sensitive information, identified by the first machine learned model, contained within the query image. Rather, the search results would relate to the rest of the query image.

The first layer of the machine learned model can be trained using a set of training data collected from a plurality of sources. In one particular embodiment, the set of training data can be images of human faces, thus causing the first machine learned model to detect regions of the query image containing human faces and eliminating them from impacting the search results. In another embodiment, the training data can be images of children or minors, thus causing the first machine learned model to detect regions of the query image containing minors and eliminating them from impacting the search results.

In some implementations, a query image can undergo a filtering step. The filtering step can be utilized to reduce unwanted search results pertaining to insignificantly small details of an image. For example, if a query image contains a high-quality image of a person dancing in some grass wearing a bright shirt that contrasts the grass, the machine learned model might return results based on a detected small region containing the contrasting colors of the shirt and grass because the contrast in pixel values might constitute a potential feature to the machine learned model. Therefore, filtering detected regions below a threshold size can help improve accuracy and relevancy of generated search results.

In some implementations, a query image is broken down into a plurality of regions. In one particular embodiment, the image is broken down into regions by dividing the image into a SxS size grid. It should be appreciated that S can be any number for generating a grid. Each piece of the grid is then passed to a first layer of a machine learned model and given a label whether or not it contains sensitive information such as, for example, a person's face. The labels can be, for example, a "1" for containing sensitive information and a "0" for not containing sensitive information. It should also be appreciated that the labels can be any form of indicator to separate one option from another and the use of "1" and "0" is for example purposes only. The regions of the grid that are given 1's are not passed to a second layer of the machine learned model due to their containing sensitive information. The regions of the grid labeled with a 0 are then passed to the second layer of the machine learned model to generate search results. In some implementations, the grid regions labeled as "1" can be further broken down into a SxS grid and re-input to the machine learned model.

In one particular embodiment, a query image is generated on a user computing device and the query image is sent to a remote computing system to perform analysis of the query image and generate search results. The search results generated by the remote computing system are then sent back to the user computing device for the user to view. In some embodiments, the query image is generated on the user computing device and a machine learned model stored within the user computing device is utilized to detect sensitive information such as, for example, a person's face. The sensitive information is then removed from the query image data and sent to a remote computing system for further processing. The query image is then input to a second machine learned model at the remote computing system and analyzed for non-sensitive information. If non-sensitive information is found, the remote computing system generates

search results and sends the search results back to the user computing device. It should be appreciated that the systems described to generate search are for descriptive purposes only. Any combination or order of the methods described herein can be executed on a user computing device, remote computing device, or similar. For example, all steps of generating search results for a query image can be performed on a remote computing system or parts of the process can be performed on a user computing device and others on a remote computing system as previously described.

Detailed Description

Figure 1 depicts an example query image process using a given query image 102. A first layer of the machine learned model 101 is used to determine the sensitive information 106 depicted in the query image 102. The sensitive information 106 can be, for example, a person's face as shown in FIG. 1, however it should be appreciated that the sensitive information 106 can be whatever the machine learned model 101 is trained to detect (e.g. minors, people, faces, etc.). The machine learned model 101 is also used to determine points of interest for search results 108 such as, for example, the building 104. The points of interests for search results 108 can by any object or area of a query image 102 that does not contain sensitive information such as, for example, the child's face 106. The query image 102 is then input to a second layer of the machine learned model 101 to generate search results 108 based on the non-sensitive information 104 within the query image 102.

Figure 2 depicts an example query image 200 being divided into regions 202 according to systems and methods of the present disclosure. It should be appreciated that a query image 200 can be broken into any number of regions 202, the grid provided in FIG. 2 is merely an example illustration for descriptive purposes. Within the regions 202 of the query image 200, there is a region 204 containing sensitive information due to the person's head being contained within the

region. In some implementations, when the query image 200 is input into a machine learned model based on aspects of the present disclosure, for example, the machine learned model 101 from FIG. 1, region 204 would be filtered from generating search results. The query image 200 also contains a region 206 that would not be deemed sensitive information because it does not contain the face of a person. In another implementation, the machine learned model is trained on persons in their entirety and therefore both regions 204 and 206 would be labeled as containing sensitive information and therefore filtered from generating search results. It should be appreciated that the examples provided herein are just a few implementations of all the possible implementations. The machine learned model can be trained to filter out anything deemed as sensitive information.

Referring now to Figure 3, an example image search process 300 is illustrated. The process comprises a query image 302 divided into a plurality of regions 303, a machine learned model 308, an intermediary image 304, and search results based on the output of the machine learned model 308, 306. The image 302 is divided into regions 303 and each region is then input into the machine learned model 308. A first layer of the machine learned model 308 generates an intermediary image 304 wherein each region 303 of the query image 302 is labeled based on whether or not it contains sensitive information. In some embodiments, the sensitive information can be, for example, a person's face, however the sensitive information may also comprise whatever the machine learned model 308 is trained to identify. The intermediary image 304 contains all the regions labeled as not containing sensitive information. The intermediary image 304 is then input to a second layer of the machine learned model 308. The region 310 is labeled as containing sensitive information by the first layer of the machine learned model 308, and therefore is not sent to the second layer of the machine learned model. In one particular

embodiment, the region containing sensitive information 310, is redivided into smaller regions (not pictured) and sent back into the first layer of the machine learned model 308.

Referring back to FIG. 3, the second layer of the machine learned model 308 checks the intermediary image 304 to determine if the remaining regions containing significant non-sensitive information such as, for example, landmarks, clothes, food, etc.. The second layer of the machine learned model 308 then generates a plurality of search results 306 based on the significant non-sensitive information found in the intermediary image 304. If no information is found in the intermediary image 304, then no results are returned. Similarly, after the first layer of the machine learned model, if all regions are found to contain sensitive information, no results are returned.

Figure 4 depicts a computer-implemented method 400 carried out by a system according to aspects of the present disclosure. At 410, a query image is broken into regions for process by a machine learned model. At 412, the regions are then provided to the machine learned model. At 414, the regions of the query image are filtered such that unusually small regions are eliminated to reduce inaccurate search results. At 416, the regions of the query image are labeled as containing or not containing sensitive information by the machine learned model. The regions containing sensitive information are then eliminated from the process. At 418, the machine learned model determines whether the remaining regions contain non-sensitive information such as, for example landmarks, objects, clothes, etc. At 420, the confidence of the detections made by the machine learned model is checked. If the confidence value is above a threshold, the output of the machine learned model generates search results relating to the determined non-sensitive information. If the confidence level is below a given threshold, no results are returned.

In some implementations, portions of the method 400 are executed on a user computing device and then the remaining portion is performed on a remote computing system. For example, steps 410-416 can be performed on a user computing device and then the results are sent to a remote computing device where steps 418-420 are performed. The results are then sent back to the user computing device for a user to view. It should be appreciated that any combination or order of the method 400 can be performed on a user computing device, remote computing device, or similar. For example, all steps of the method 400 can be performed on a user computing device or a remote computing system.

Figures

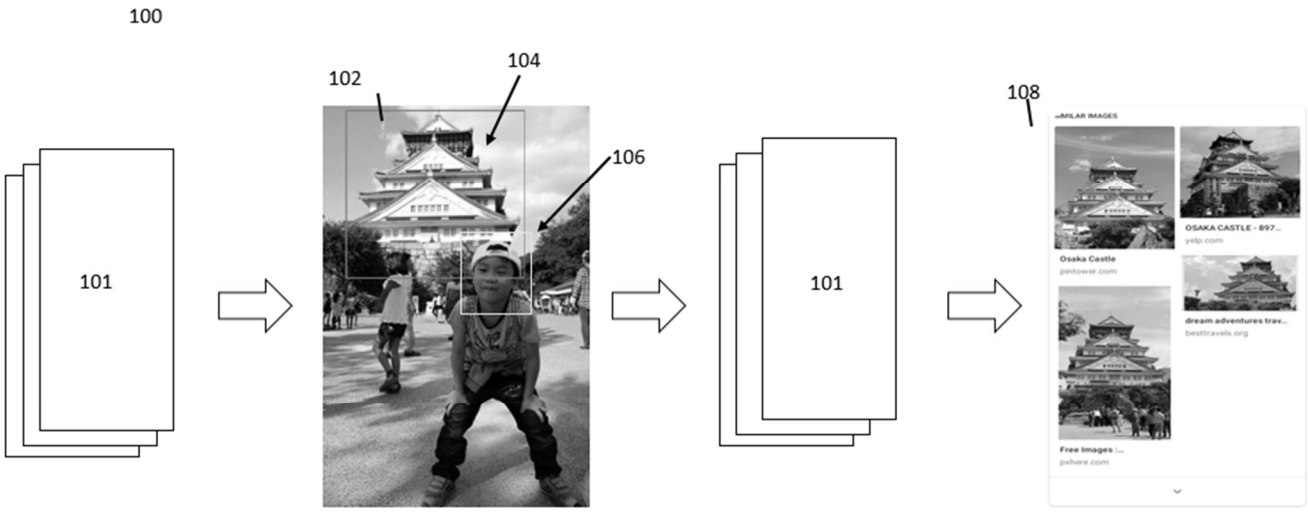

FIG. 1

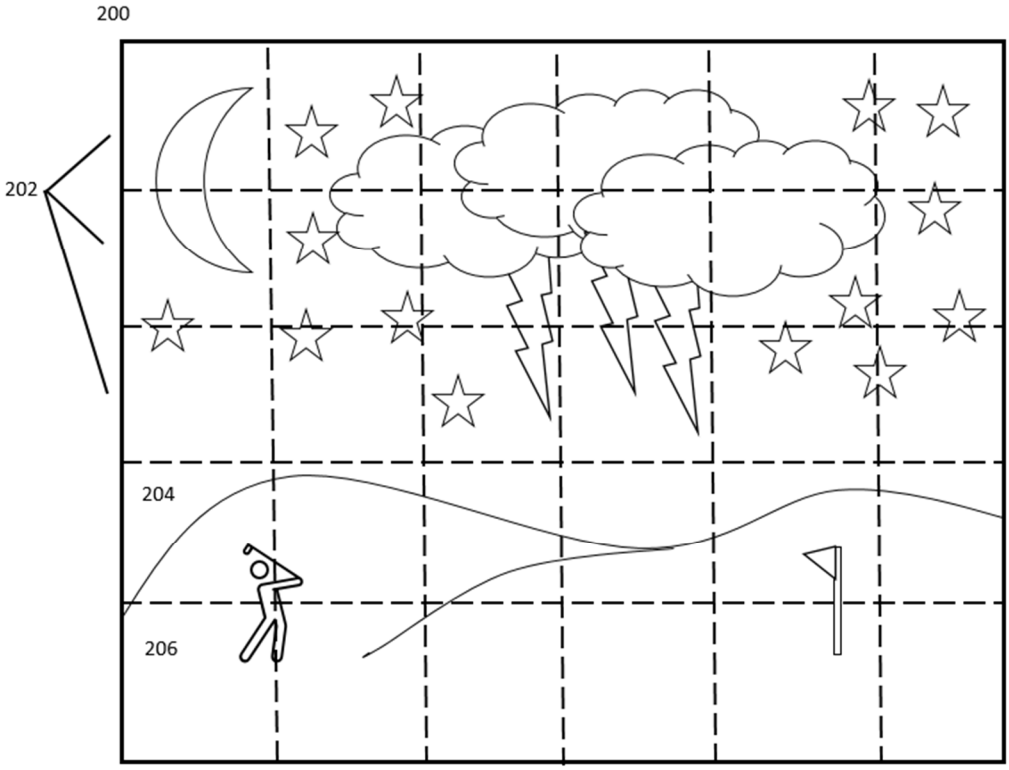

FIG. 2

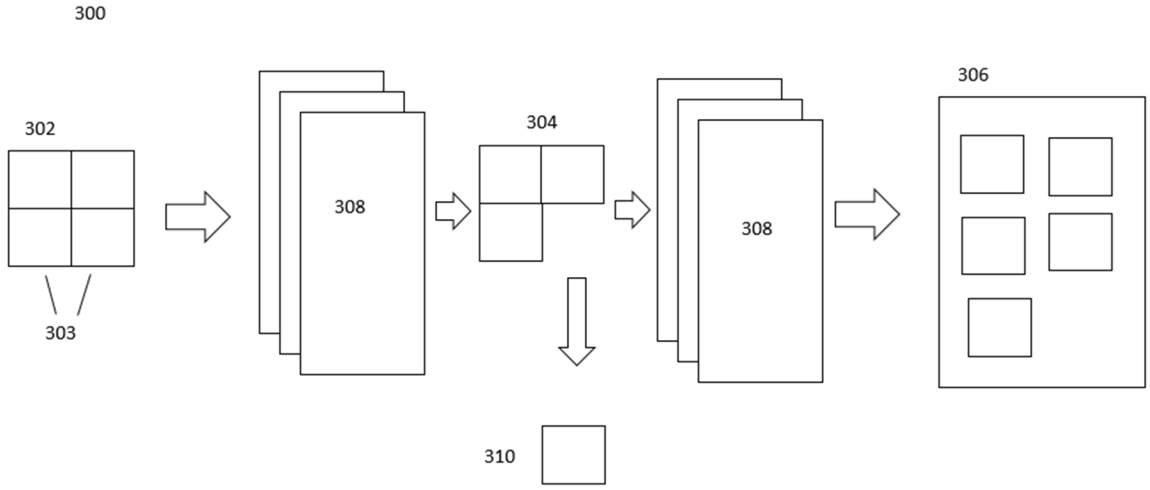

FIG. 3

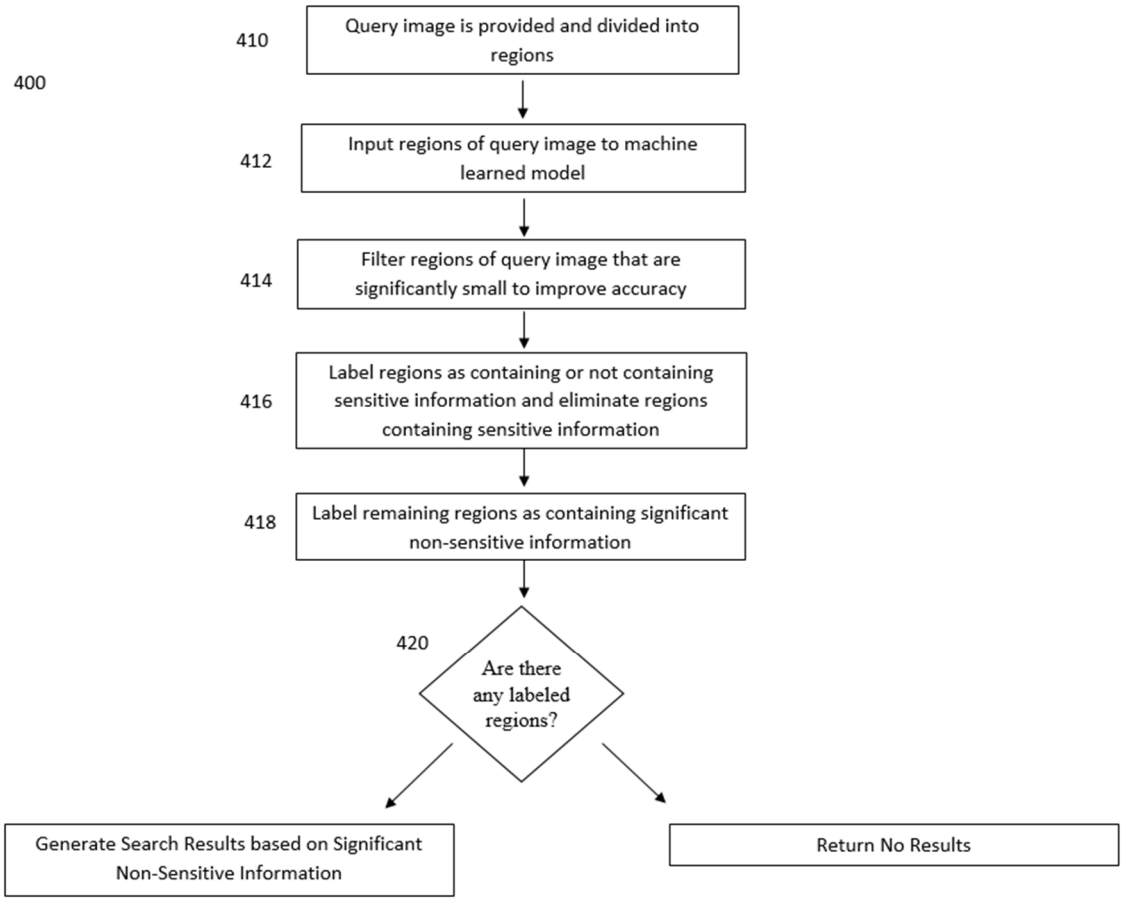

FIG. 4

Abstract

The present disclosure describes computer-implemented systems and methods for generating search results from a user provided query image that may contain or contains sensitive information (e.g. a minor's face). When the system is provided a query image, the system analyzes the image for potentially sensitive information such as, for example, a person's face and filters it out. The system then searches the remaining image for potential search results relating to information detected within the image (objects, places, clothes, etc.).