(10) Publication No: **IN586374 B1**

(45) Date of Grant: 06-04-2026

(19) Intellectual Property Office, India

(21) Application No: 202141032307

(24) Date of Patent: 19-07-2021

(12) INDIAN GRANTED PATENT

(54) Title: **A system and method for searching an object in a target image using augmentation**

(51) International Classification:
G06K9/62, G06N3/04, G06N3/08, G06K9/00, G06K9/46

(22) Date of Filing: **19-07-2021**

(43) Publication Date: **20-01-2023**
Journal No: **03/2023**

(31) Priority Document No: **--**

(32) Priority Date: **--**

(86) International Application No: **--**
Filing Date: **--**

(87) International Publication No: **--**

(61) Patent of Addition to Application No: **--**
Filing Date: **--**

(62) Divisional to Application No: **--**
Filing Date: **--**

(71) Name of Applicant(s):
1. Robert Bosch Engineering And Business Solutions Private Limited,
2. Robert Bosch Gmbh,

(72) Name of Inventor(s):
1. Amit Arvind Kale
2. Sonam Singh

(57) Abstract:

A SYSTEM AND METHOD FOR SEARCHING AN OBJECT IN A TARGET IMAGE USING AUGMENTATION ABSTRACT Method (200) and system (400) for searching an object in an image (102), are described. An image (102) of the object is provided as an input to a convolutional neural network (CNN) (300). A processor (112) of the computing system (400) provides an image (102) of the object as an input query to a pre-trained convolutional neural network (CNN) (302). The pre-trained convolutional neural network (CNN) (302) performs augmentation on the input image (102), by applying semantic preserving transformations and obtain a set of augmented images (102-1 to 102-N) representing the input image (102). The CNN 302 further configured to extract plurality of feature vectors (108) from each of obtained augmented images (102-1 to 102-N) indicative of the input image (102), and merge the extracted feature vectors (108) together to generate a single feature vector (110) representing the input image (102). Based on the merged feature vector (116), retrieving the search results representing search object in the image (102). (Figure 1)

FORM 2

THE PATENTS ACT, 1970
(39 of 1970) & The Patents Rules 2003

COMPLETE SPECIFICATION
(SECTION 10 and Rule 13)

1. **Title of the Invention: A system and method for searching an object in a target image using augmentation**

2. **Applicants:**

a. Name: Robert Bosch Engineering and Business Solutions Private Limited
Nationality: INDIA
Address: 123, Industrial Layout, Hosur Road, Koramangala, Bangalore 560095, Karnataka, India

b. Name: Robert Bosch GmbH
Nationality: GERMANY
Address: Feuerbach, Stuttgart, Germany

Complete Specification:

The following specification describes and ascertains the nature of this invention and the manner in which it is to be performed

Field of the invention

[0001] The present subject matter relates, in general, to image searching, and, particularly, to searching an object in a target image using augmentation.

Background of the invention

[0002] Various image retrieval techniques involve computer-based object recognition. The computer-based object recognition incorporates image processing to identify instances of objects in images and performs annotations of the objects for object detection and tracking. Generally, in keywords based image retrieval system it can be difficult or impossible for a user to precisely describe the inherent complexity of certain images. As a result, retrieval accuracy can be severely limited because images that cannot be described or can only be described ambiguously will not be retrieved successfully.

[0003] Content based image retrieval (CBIR) systems finds images which are similar to low level image feature of an example image, such as color histogram, texture, shape, and so forth. Although CBIR solves the problem of keyword-based image retrieval, it also has severe shortcomings. However, in such CBIR systems, the searches may return entirely irrelevant images that just happen to possess similar features. Additionally, individual objects in images contain a wide variety of low-level features. Therefore, using only the low level features will not satisfactorily describe what is to be retrieved. Image retrieval systems must understand users 'intent while processing their queries. This is a difficult task which become more severe in case of image queries due to inherent ambiguity caused by the absence of semantic information, textures, colors, and luminance.

[0004] The prior art, US8433705B1 discloses a method searches a set of information using a computer. The method generates a set of search results based on a search query. Then, without further user input, the method generates a set of candidate facets, where each of the candidate facets can be used to select a subset of the search results. The method ranks the candidate facets in accordance with selectivity of the candidate facets and selects a plurality of facets from among the candidate facets for presentation to the user. The selection is in accordance with the rankings of the candidate facets. The method formats the presentation facets for display to the user. In response to user selection of a presentation facet, the method generates a revised search query comprising the original search query and the selected presentation facet, and generates a revised set of search results based on the revised search query.

[0005] Another prior art US7472113B1 discloses Front end preprocessing modifies queries to get them in better form for presentation to a variety of data sources. The queries are modified by grouping terms as phrases, correcting spelling errors, and augmenting the query with category terms that trigger query execution on certain data sources. Context information about a particular user as well as information about previous queries by other uses are also used to modify the query to better reflect the user's intent.

Brief description of the accompanying drawing

[0006] The detailed description is provided with reference to the accompanying figures, wherein:

[0007] FIG. 1 illustrates a system environment for image searching, in accordance with an example implementation of the present subject matter; and

[0008] FIGS. 2 illustrates a flow chart of a method for image searching, in accordance with an example implementation of the present subject matter.

Detailed description of the embodiments

[0009] FIG. 1 illustrates a system environment for image searching, in accordance with an example implementation of the present subject matter. The present subject matter describes various approaches for improving search for a particular object represented by a query image using augmentation and images containing similar objects, are retrieved. In an example implementation of the present subject matter, a method and a system are proposed for improving search for a particular object represented by a query image.

[0010] The system environment may include a computing system 400 and a neural network architecture. The computing system 400 may be communicatively coupled to the neural network architecture. In an example, the computing system 400 may be directly or remotely coupled to the neural network architecture. Examples of the computing system 400 may include, but are not limited to, a laptop, a notebook computer, a desktop computer, and so on.

[0011] The present subject matter involves providing an image 102 of the object as an input to a pre-trained neural network, such as a pre-trained convolutional neural network (CNN). The CNN may be capable of extracting features from the image. In an example, the CNN comprises a layered structure having a set of convolutional layers and pooling units. The set of convolutional layers and pooling units are arranged to facilitate in handling of objects of different dimensions in an input image. The set of convolutional layers and pooling units may receive the input and convolute the input with one or more image filters.

[0012] The neural network architecture may be a convolutional neural network (CNN) architecture 300. The CNN architecture 300 may include a pre-trained CNN 302 (hereinafter referred to as “CNN 302”) and a CNN-based feature database 304. The CNN 302 may represent a machine learning model including a set of input layers, a set of hidden layers, and a set of output layers. The CNN 302 may be a pre-trained CNN which may use a pre-existing model, trained on a large dataset of images, for feature extraction. Further, the CNN-based feature database 304 may include data pertaining to the CNN 302, such as learning techniques, deep recognition patterns, and so on. For example, the CNN-based feature database 304 may store a plurality of feature vectors pre-extracted from a plurality of images.

[0013] The CNN may also include a CNN-based feature database in communication with a library of images. Each image from the library of image may include one or more instances of the object having a wide variance of appearance. The variation of appearance may involve different orientations, illumination conditions, image scale, image quality, and so on. The image may include the object, which is required to be searched, along with multiple other objects of different classes. In an example, the CNN-based feature database may include a pre-fetched plurality of feature vectors associated with the images included in the library of images.

[0014] The computing system 400 may include a memory 110. The memory 110 may include any non-transitory computer-readable medium including, for example, volatile memory, such as static random-access memory (SRAM) and dynamic random-access memory (DRAM), and/or non-volatile memory, such as read only memory (ROM), erasable programmable ROM, flash memories, hard disks, optical disks, and magnetic tapes.

[0015] In an example, the computing system 400 may also include a processor 112 coupled to the memory 110. The processor 112 may include microprocessors, microcomputers, microcontrollers, digital signal processors, central processing units, state machines, logic circuitries, and/or any other devices that manipulate signals and data based on computer-readable instructions. Further, functions of the various elements shown in the figures, including any functional blocks labelled as "processor(s)", may be provided through the use of dedicated hardware as well as hardware capable of executing computer-readable instructions.

[0016] Further, the computing system 400 may include interface(s) 114. The interface(s) 114 may include a variety of interfaces, for example, interface(s) for users. The interface(s) 114 may include data output devices. In an example, the interface(s)

114 may provide an interactive platform for receiving inputs from a user. For example, the user may provide an image 102 of an object as an input to the computing system 400 through the interface(s) 114. The image provided by a user may also be referred as an input image. In an example, the image may be an image of a traffic sign, vehicle(s), pedestrian(s), and so on.

[0017] In one embodiment, the processor 112 may receive the input image 102 of the object from a user. In an example, the user may upload the input image 102 to the computing system 400 as a query. Further, the image of the object (hereinafter input image 102) is provided to a pre-trained convolutional neural network (CNN) 302.

[0018] The processor 112 configured to perform augmentation techniques on the input image 102 provided, by applying semantic preserving transformations. Further, a set of augmented images (102-1 to 102-N) representing the input image 102, are obtained after performing the augmentation on the input image 102. In an exemplary embodiment, the semantic preserving transformations comprises but are not limited to color jitter, blurring, rotation and the like.

[0019] Further, the set of augmented images (102-1 to 102-N) representing the input image 102 may be provided, by the processor 112, to the CNN 302 as an input for extraction of plurality of features pertaining to the input image. A plurality of feature vectors 108 from each of obtained augmented images (102-1 to 102-N) indicative of the input image 102, are extracted from the pre-trained CNN 302. The feature may represent a set of characteristics, such as a shape, color profiles, texture pattern, or a combination thereof, associated with the input image. In an example, the CNN 302 may generate feature vector from the set of features. The feature vector may be a representation of the input image 102. The extracted feature vectors 108 are merged by the CNN 302, together to generate a single feature vector 116 representing the input image 102 provided as an input query. The merged feature vector 116 encodes the invariance through the augmented versions of the query.

[0020] In one embodiment of the present invention, the processor 112 may compare the merged feature vector 116 with the plurality of feature vectors corresponding to a list of images stored in the CNN-based feature database 304. In an example, the plurality of feature vectors may be pre-extracted from the plurality of images of the list of images. In an example, the CNN-based feature database 304 may be in communication with a pre-indexed library of images 306 to extract the list of images. In another example, the CNN-based feature database 304 may include the pre-indexed library of images 306.

[0021] In an example, the processor 112 may compute a similarity score of the merged feature vector 116 with each of the plurality of feature vectors stored in the CNN-based feature database 304. The similarity score may be computed by comparing the merged feature vector 116 with each of the plurality of feature vectors stored in the CNN-based feature database 304.

[0022] Further, based on the similarity score, the processor 112 may obtain one or more target images, each of the one or more target images having a similarity score above a predetermined threshold. The predetermined threshold may be based on a determination of presence of at least the object, present in the input image, in an image from the list of images. In an example, the input image may form a subset of each of the one or more target images and the target image may contain features associated with the input image. For example, if the input image includes a pedestrian as an object, a target image may include at least a pedestrian having features similar to the pedestrian present in the input image.

[0023] Further, the processor 112 may retrieve the search results representing search object in the image 102, from the CNN 302, based on the merged feature vector 116. In one embodiment, the search results representing search object in the image 102 may include images containing the similar objects. In one embodiment, the retrieved search results may be presented to the user via the computing system 400 through the interface 114. In this embodiment, the user may download a subset of list of images relevant for his/her task. Further in this embodiment, the user can select one of the relevant result image from the retrieved search results, to start the search process again by drawing bounding box around an object of interest in the selected relevant image.

[0024] FIG. 2 illustrates a flow chart of a method 200 for image searching in accordance with an example implementation of the present subject matter. The method 200 may be implemented by the computing system 400 including the memory 110, the processor 112, and the interface(s) 114, of FIG. 1. Further, the computing system 400 may be communicatively coupled with the neural network architecture as described in FIG. 1. The neural network architecture may include a CNN 302 and a CNN-based feature database 304. Although, the method 200 is described in context of the system that is similar to the computing system 400 of FIG. 1, other suitable devices or systems may be used for execution of the method 200.

[0025] Referring to FIG.2, an input image 102 of an object may be provided by the user as a query to the computing system 400. The image may be provided by the user through the interface(s) 114. The image provided by the user may also be referred hereinafter as an input image. Further, the input image may be provided as an input to the CNN 302 for extraction of features pertaining to the input image. The set of features may be a representation of a set of characteristics of the input image. For example, the set of characteristics may include a shape, color profiles, texture pattern, or a combination thereof.

[0026] At block 201, the method 200 may include providing, by a processor 112 of a computing system 400, an image 102 of the object as an input query to a pre-trained convolutional neural network CNN 302. At block 202, the method 200 may include performing augmentation on the input image 102 provided as an input query by applying semantic preserving transformations and obtaining a set of augmented images (102-1 to 102-N) representing the input image 102 provided as an input query. In an exemplary embodiment, the semantic preserving transformations comprises but are not limited to color jitter, blurring, rotation and the like.

[0027] At block 203, the method 200 may include extracting plurality of feature vectors 108 from each of obtained augmented images (102-1 to 102-N) indicative of the input image 102, from the pre-trained CNN 302. The feature may represent a set of characteristics, such as a shape, color profiles, texture pattern, or a combination thereof, associated with the input image. In an example, the CNN 302 may generate feature vector from the set of features. The feature vector may be a representation of the input image 102.

[0028] In addition, at block 204, the method 200 may include, merging the extracted feature vectors 108 together to generate a single feature vector 110 representing the input image 102 provided as an input query. The merged feature vector 116 encodes the invariance through the augmented versions of the query.

[0029] In one embodiment of the present invention, the processor 112 may compare the merged feature vector 116 with the plurality of feature vectors corresponding to a list of images stored in the CNN-based feature database 304. In an example, the processor 112 may compute a similarity score of the merged feature vector 116 with each of the plurality of feature vectors stored in the CNN-based feature database 304. For example, the similarity score may define a degree of relevance of each of the images present in the CNN-based feature database 304. In an example, the list of images may be retrieved from the CNN-based feature database 304 based on the similarity score. Further, at block 205, the method 200 may include, based on the merged feature vector 116, retrieving the search results representing search object in the image 102.

[0030] Although aspects for the present disclosure have been described in a language specific to structural features and/or methods, it is to be understood that the appended claims are not limited to the specific features or methods described herein.

Rather, the specific features and methods are disclosed as examples of the present disclosure.

Clean copy of claims

CLAIMS

We Claim:

1. A method (200) for searching an object in an image (102), the method (200) comprising the steps for:

providing (201), by a processor (112) of a computing system (400), an image (102) of the object as an input query to a pre-trained convolutional neural network (CNN) (302);

performing (202) augmentation on the input image (102) provided as an input query by applying semantic preserving transformations and obtaining a set of augmented images (102-1 to 102-N) representing the input image (102) provided as an input query;

Characterized in that :

extracting (203) plurality of feature vectors (108) from each of obtained augmented images (102-1 to 102-N) indicative of the input image (102), from the pre-trained CNN (302);

merging (204) the extracted feature vectors (108) together to generate a single feature vector (110) representing the input image (102) provided as an input query;

based on the merged feature vector (116), retrieving (205) the search results representing search object in the image (102).

2. The method (200) as claimed in claim 1, wherein the method (200) comprising a step for comparing the merged feature vector (116) with a plurality of feature vectors corresponding to a list of images stored in a CNN-based feature database (304).

3. The method (200) as claimed in claim 2, wherein comparing the merged feature vector (116) comprises computing a similarity score of the merged

feature vector (116) with each of the plurality of feature vectors stored in the CNN-based feature database (304).

4. The method (200) as claimed in claim 1, wherein the search results representing search object in the image (102) comprises images containing the similar objects.

5. The method (200) as claimed in claim 1, wherein examples of the semantic preserving transformations comprises but are not limited to color jitter, blurring, rotation and the like.

6. The method (200) as claimed in claim 1, wherein the merged feature vector (116) encodes the invariance through the augmented versions of the query.

7. The method (200) as claimed in claim 1, wherein said retrieved search results are presented to the user via the computing system (400) through the interface (114).

8. The method (200) as claimed in claim 7, wherein said user selects one of the relevant result image from the retrieved search results, to start the search process again by drawing bounding box around an object of interest in the selected relevant image.

9. A computing system (400) comprising:
a memory (110); and
a processor (112), coupled to the memory (110), to:
provide an image (102) of the object as an input query to a pre-trained convolutional neural network (CNN) (302);
perform augmentation on the input image (102) provided as an input query by applying semantic preserving transformations;

obtain a set of augmented images (102-1 to 102-N) representing the input image (102) provided as an input query;

characterized in that :

extract plurality of feature vectors (108) from each of obtained augmented images (102-1 to 102-N) indicative of the input image (102), from the pre-trained CNN (302);

merge the extracted feature vectors together to generate a single feature vector (116) representing the input image (102) provided as an input query;

retrieve the search results representing search object in the image (102), based on based on the merged feature vector (116).

Dated this 7th day of Sep 2023

(Digitally signed)
Siddharth Karkhanis
On-behalf of the Applicants (IN/PA-1195)

A SYSTEM AND METHOD FOR SEARCHING AN OBJECT IN A TARGET IMAGE USING AUGMENTATION

ABSTRACT

Method (200) and system (400) for searching an object in an image (102), are described. An image (102) of the object is provided as an input to a convolutional neural network (CNN) (300). A processor (112) of the computing system (400) provides an image (102) of the object as an input query to a pre-trained convolutional neural network (CNN) (302). The pre-trained convolutional neural network (CNN) (302) performs augmentation on the input image (102), by applying semantic preserving transformations and obtain a set of augmented images (102-1 to 102-N) representing the input image (102). The CNN 302 further configured to extract plurality of feature vectors (108) from each of obtained augmented images (102-1 to 102-N) indicative of the input image (102), and merge the extracted feature vectors (108) together to generate a single feature vector (110) representing the input image (102). Based on the merged feature vector (116), retrieving the search results representing search object in the image (102).

(Figure 1)

Applicant: 1. Robert Bosch Engineering and Business Solutions Private Ltd — Sheet 1 of 2
2. Robert Bosch GmbH — Total Sheets 2

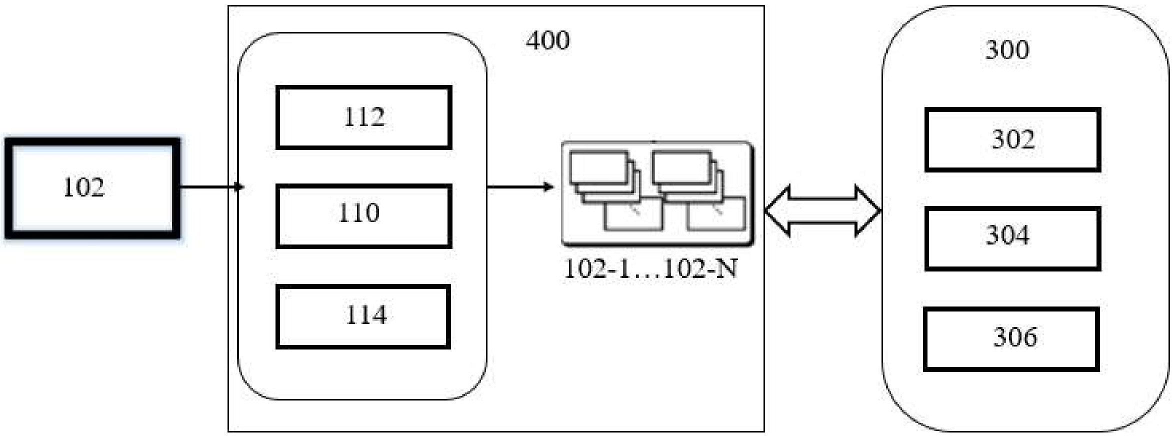

Figure 1

(Digitally signed)
Siddharth Karkhanis
On-behalf of the Applicants (IN/PA-1195)

Applicant: 1. Robert Bosch Engineering and Business Solutions Private Ltd 2. Robert Bosch GmbH

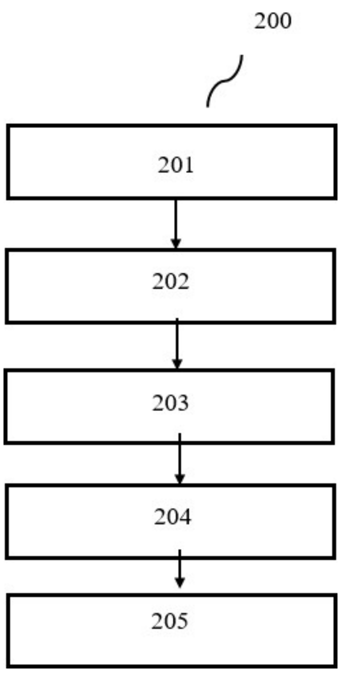

Figure 2

(Digitally signed)
Siddharth Karkhanis
On-behalf of the Applicants (IN/PA-1195)