



(51) International Patent Classification:

G06N 20/00 (2019.01) G06K 9/62 (2006.01)
G06N 3/02 (2006.01) G06T 7/33 (2017.01)

(21) International Application Number:

PCT/SG2020/050642

(22) International Filing Date:

06 November 2020 (06.11.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: **VISENZE PTE LTD** [SG/SG]; 67 Ayer Rajah Crescent, #07-05/09, Singapore 139950 (SG).

(72) Inventors: **LI, Guangda**; 5 Dairy Farm Heights, #05-11, The Skywoods, Singapore 677668 (SG). **CAO, Thanh Tung, Todd**; 67 Ayer Rajah Crescent, #07-05/09, Singapore 139950 (SG). **WANG, Zhenhua**; 156 Haig Road, Haig Court, #09-06, Singapore 438793 (SG). **Ji, Xin**; 460 Clementi Avenue 3, #05-600, Singapore 120460 (SG).

(74) Agent: **ORION IP LLP**; 1 Sophia Road, Peace Centre, #07-21, Singapore 228149 (SG).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: A SYSTEM AND A METHOD FOR GENERATING AN IMAGE RECOGNITION MODEL AND CLASSIFYING AN INPUT IMAGE

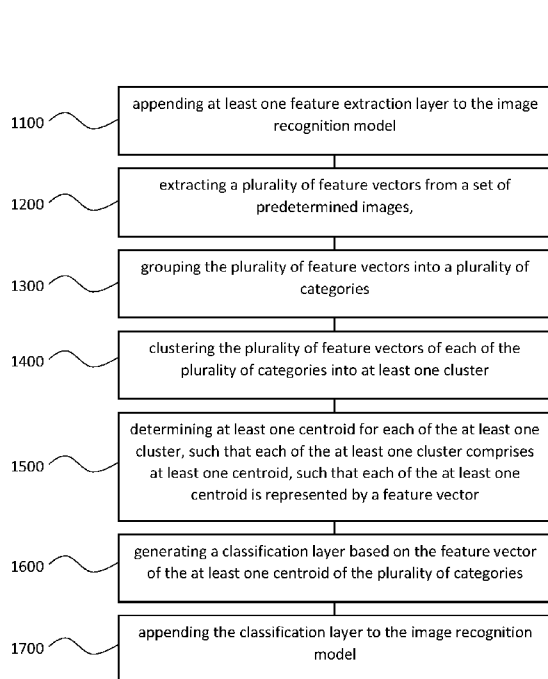


Fig. 1

(57) Abstract: A method of generating an image recognition model for recognising an input image and a system thereof are provided. The method includes appending at least one feature extraction layer to the image recognition model, extracting a plurality of feature vectors from a set of predetermined images, grouping the plurality of feature vectors into a plurality of categories, clustering the plurality of feature vectors of each of the plurality of categories into at least one cluster, determining at least one centroid for each of the at least one cluster, such that each of the at least one cluster comprises at least one centroid, such that each of the at least one centroid is represented by a feature vector, generating a classification layer based on the feature vector of the at least one centroid of the plurality of categories, and appending the classification layer to the image recognition model. In addition, a method of classifying an input image and a system thereof are provided.



A SYSTEM AND A METHOD FOR GENERATING AN IMAGE RECOGNITION MODEL AND CLASSIFYING AN INPUT IMAGE

Technical Field

[0001] The present invention relates to a system and a method for generating an image recognition model for recognising an input image. The present invention further relates to a system and a method for classifying an input image.

Background

[0002] Due to the remarkable image recognition performance, deep learning methods have been widely researched and applied. A traditional deep learning image recognition model consists of a feature extraction module and a classification module. The recognition process using such a model includes taking an image of a given object and inputting it to the feature extraction module, which consists of a number of neural network layers, to generate a feature vector, basically a list of numbers, for the image. Subsequently, the classification module is used to process the feature vector and identify the object by categorizing it into one of the predefined classes or identities. Generally, a classification module consists of one classification neural network layer, which is typically a fully connected layer of nodes. When recognizing an image, the classification module applies this layer on the feature vector obtained from the feature extraction module, and produces a list of numbers, each corresponds to a predefined category, and the actual value represents the similarity between the image and the category. The larger the value is, the more similar is the image to the category. Finally, the name of the category with maximum value will be retrieved and returned and the image is thus recognized. When using the above deep learning image recognition model, the entire recognition process is calculated end-to-end in a neural network model. The whole model can be trained end-to-end using a deep learning algorithm such as back-propagation, given sufficient training data. Such an approach has a simple structure, and has been greatly optimized for computational efficiency. However, one major disadvantage is that the classification layer in the classification module is fixed after training. Once the model was trained, it is difficult to either add one more category for the model to recognize, or tune the performance of one specific category, without

re-training the whole model again, which will often be time-consuming and affect the model's performance on all categories.

[0003] A search-based recognition model, on the other hand, is much more flexible than a classification model with regard to the above-mentioned aspects. A search-based recognition model has no classification layer in its classification module. Instead, it consists of a database of a large number of indexed feature vectors of a set of images. These feature vectors are pre-extracted by the feature extraction module. To recognise an input image, after receiving the input feature vector extracted from the feature extraction module, the recognition module will scan its database and retrieve the indexed feature vector nearest (i.e. most similar) to that input feature vector, and return the category name associated with that indexed feature vector. Compared to the conventional image recognition model mentioned in the previous paragraph, such a search-based recognition model is much more flexible with the introduction of the database. By adding, removing or updating the feature vectors in the database in the recognition module, it is possible not only to improve the image recognition performance of an existing category, but also to modify the list of categories recognizable by the model. In other words, it is now possible to fine-tune such a model without re-training the underlying neural network layers.

[0004] However, a search-based recognition model also has its disadvantages. The feature database size is often large since it is proportional to the numbers of images added, which is typically the whole training set, instead of the number of categories. As a result, it often consumes more memory resources than a traditional recognition model, and it also requires more computation when performing a recognition task. Therefore, it is more difficult to deploy the model on a system with limited resources, e.g. a mobile device. It is possible to reduce the database size by only indexing a small number of images, even just one image, for each category, but the process of selecting the right images to index in order to preserve the recognition performance is non-trivial.

[0005] It is therefore necessary to have a new type of recognition model that is able to overcome the above-mentioned issues. Preferably, it is structurally simple, easy to update and improve, while still able to achieve a good recognition accuracy.

Summary

[0006] According to various embodiments, a method of generating an image recognition model for recognising an input image is provided. The method includes appending at least one feature extraction layer to the image recognition model, extracting a plurality of feature vectors from a set of predetermined images, grouping the plurality of feature vectors into a plurality of categories, clustering the plurality of feature vectors of each of the plurality of categories into at least one cluster, determining at least one centroid for each of the at least one cluster, such that each of the at least one cluster includes at least one centroid, such that each of the at least one centroid is represented by a feature vector, generating a classification layer based on the feature vector of the at least one centroid of the plurality of categories, and appending the classification layer to the image recognition model.

[0007] According to various embodiments, the method may include varying the number of clusters of each of the plurality of categories into a new number of clusters, re-clustering the plurality of feature vectors of each of the plurality of categories into the new number of clusters, re-determining at least one centroid for each of the new number of clusters wherein each of the at least one centroid is represented by a feature vector, and re-generating the classification layer of the image recognition model based on the feature vector of the at least one centroid of each of the new number of clusters.

[0008] According to various embodiments, the method may include receiving a plurality of images for a new category, extracting a plurality of new feature vectors from the plurality of images by the feature extraction module, clustering the plurality of new feature vectors into at least one new cluster, generating at least one centroid for each of the at least one new cluster, wherein each of the at least one centroid is represented by a feature vector, and appending the feature vector of each of the at least one centroid for each of the at least one new cluster to the classification layer of the image recognition model.

[0009] According to various embodiments, the method may include receiving a plurality of selected images in one of the plurality of categories, extracting a plurality of new feature vectors of the plurality of selected images, adding the plurality of new feature vectors to the

plurality of feature vectors of the one of the plurality of categories to form a new set of feature vectors, re-clustering the new set of feature vectors into at least one cluster, re-generating at least one centroid for each of the at least one cluster, such that each of the at least one centroid is represented by a feature vector, and appending the feature vector of each of the at least one centroid for each of the at least one cluster to the classification layer of the image recognition model.

[0010] According to various embodiments, the plurality of feature vectors may be clustered using the k-mean algorithm.

[0011] According to various embodiments, a method for classifying an input image is provided. The method includes receiving the input image, inputting the image to the generated image recognition model as mentioned above to classify the input image into one of the plurality of categories, and outputting a category name of the one of the plurality of categories of the input image from the image recognition model.

[0012] According to various embodiments, the method may include extracting a feature vector of the input image, comparing the feature vector to the plurality of centroids in the classification layer, determining one of the plurality of centroids that is nearest to the feature vector; identifying the category of the one of the plurality of centroids; and retrieving a category name of the category of the one of the plurality of centroids.

[0013] According to various embodiments, comparing the feature vector to the plurality of centroids may include calculating the distances between the feature vector and each of the feature vector of the at least one centroid of the plurality of categories.

[0014] According to various embodiments, a system for generating an image recognition model for recognising an input image is provided. The system includes a processor, and a memory in communication to the processor for storing instructions executable by the processor, such that the processor is configured to append at least one feature extraction layer to an image recognition model, extract a plurality of feature vectors from a set of pre-determined images, group the plurality of feature vectors into a plurality of categories, cluster the plurality of feature vectors of each of the plurality of categories into at least one cluster, determine at least

one centroid for each of the at least one cluster, such that each of the at least one cluster comprises at least one centroid, wherein each of the at least one centroid is represented by a feature vector, generate the classification layer based on the feature vector of the at least one centroid of the plurality of categories; and append the classification layer to the image recognition model.

[0015] According to various embodiments, the processor may be configured to vary the number of clusters of each of the plurality of categories into a new number of clusters, re-cluster the plurality of feature vectors of each of the plurality of categories into the new number of clusters, re-determine at least one centroid for each of the new number of clusters wherein each of the at least one centroid is represented by a feature vector, re-generate the classification layer in the recognition model based on the feature vector of the at least one centroid of each of the new number of clusters.

[0016] According to various embodiments, the processor may be configured to receive a plurality of selected images in one of the plurality of categories, extract a plurality of new feature vectors of the plurality of selected images, add the plurality of new feature vectors to the plurality of feature vectors of one of the plurality of categories to form a new set of feature vectors, re-cluster the new set of feature vectors into at least one cluster, re-generate at least one centroid for each of the at least one cluster, such that each of the at least one centroid is represented by a feature vector, and append the feature vector of each of the at least one centroid for each of the at least one cluster to the classification layer to the image recognition model.

[0017] According to various embodiments, the plurality of feature vectors may be clustered using k-mean algorithm.

[0018] According to various embodiments, a system for classifying an input image is provided. The system includes a processor, and a memory in communication to the processor for storing instructions executable by the processor, such that the processor is configured to receive the input image, input the input image to the generated image recognition model as mentioned above to classify the input image into one of the plurality of categories; and output a category name of the one of the plurality of categories of the input image from the image recognition model.

[0019] According to various embodiments, the processor may be configured to extract a feature vector of the input image, compare the feature vector to the plurality of centroids in the classification layer, determine one of the plurality of centroids that is nearest to the feature vector, identify the category of the one of the plurality of centroids and retrieve the category name of the category of the one of the plurality of centroids.

[0020] According to various embodiments, a non-transitory computer readable storage medium comprising instructions is provided, such that the instructions, when executed by a processor in a system, cause the system to append at least one feature extraction layer to an image recognition model, extract a plurality of feature vectors from a set of pre-determined images, group the plurality of feature vectors into a plurality of categories, cluster the plurality of feature vectors of each of the plurality of categories into at least one cluster, determine at least one centroid for each of the at least one cluster, such that each of the at least one cluster comprises at least one centroid, such that each of the at least one centroid is represented by a feature vector, generate the classification layer based on the feature vector of the at least one centroid of the plurality of categories, and append the classification layer to the image recognition model.

[0021] According to various embodiments, a non-transitory computer readable storage medium comprising instructions is provided, such that the instructions, when executed by a processor in a system, cause the system to receive an input image, input the input image to the generated image recognition model as mentioned above to classify the input image into one of the plurality of categories, and output a category name of the one of the plurality of categories of the input image from the image recognition model.

Brief Description of Drawings

[0022] Fig. 1 shows an exemplary flow diagram of a method for generating an image recognition model for recognising an input image.

[0023] Fig. 2 shows an exemplary flow diagram of the feature extraction model recognising an input image.

[0024] Fig. 3 shows an exemplary flow diagram of the image recognition model classifying an input image.

[0025] Fig. 4 shows a flow diagram of a method for classifying an image.

[0026] Fig. 5 shows an exemplary flow diagram of the image recognition model identifying an input image.

[0027] Fig. 6 shows a schematic view of an exemplary embodiment of a system for deploying the methods in Fig. 1.

Detailed Description

[0028] Fig. 1 shows an exemplary flow diagram of a method 1000 for generating an image recognition model for recognising an input image. Method includes appending at least one feature extraction layer to the image recognition model in block 1100, extracting a plurality of feature vectors from a set of predetermined images in block 1200, grouping the plurality of feature vectors into the plurality of categories in block 1300, clustering the plurality of feature vectors of each of the plurality of categories into at least one cluster in block 1400, determining at least one centroid for each of at least one cluster in block 1500, such that each of the at least one cluster includes at least one centroid, such that each of the at least one centroid is represented by a feature vector, generating a classification layer based on the feature vector of the at least one centroid of the plurality of categories in block 1600, and appending the classification layer to the image recognition model in block 1700. The method may be implemented by a system, e.g. a computer, a server, a mobile device, etc.

[0029] The image recognition model may include a feature extraction module and a classification module. The feature extraction module may include the at least one feature extraction layer. The classification module may include a classification layer. From the feature extraction module, a conversion and compression process may be carried out to generate the parameters for generating the classification layer. By having a “one class multiple centroids”

in the classification model, as will be shown later, it is possible to obtain a balanced model size and recognition accuracy. There may be a plurality of classes in the classification layer. This image recognition model has the flexibility of a search-based recognition model, while retaining the deployment simplicity and efficiency of a traditional classification model. Consequently, the image recognition model is more suitable to be deployed on a resource constrained system, such as a mobile device.

[0030] Fig. 2 shows an exemplary flow diagram of the image recognition model 230 processing an input image 210. Image recognition model 230 may be a search-based recognition model. Image recognition model 230 may include the feature extraction module 220. Image recognition model 230 may include a recognition module configured to recognise an image. Recognition module may include a feature vector database. Feature vector database may store feature vectors of images and the category name of the feature vectors. As shown in Fig. 2, the feature extraction module 220 may include at least one feature extraction layer 222. The at least one feature extraction layer 222 may include a stack of convolutional neural network layers. The layers may be weighted and weights (parameters) of these layers may be commonly known in the state of the art. The feature extraction module 220 may be pre-trained using a deep learning algorithm. A common training method may be supervised training. For example, a taxonomy of categories for the recognition target may first be defined and a set of predetermined images related to the taxonomy of categories may be used as training data to train the at least one feature extraction layer 222 using a classification objective function. When the objective function converges to below a certain threshold, the training can be stopped and the at least one feature extraction layer 222 may be used to extract feature vectors from the input image.

[0031] The feature extraction module 220 may be configured to extract feature vectors from a predetermined set of images, i.e. feature vectors of feature items of the predetermined set of images. This set of images may be the same as the images used to train the feature extraction module 220, or images in a separate search database. To recognise an input image 210, the feature extraction module 220 may convert the input image 210, which may simply be a three-dimensional float tensor, to a one-dimensional float vector, referred to as the embedding 224 of the input image 210. The embedding 224 may be matched against feature vectors in the feature database of the recognition module. When the feature vector, e.g. of a shark, is matched

to the input image embedding 224 as the most similar one, the category name associated with the category of the feature vector may be retrieved, e.g. “shark”, and returned, and the input image 210 is recognised. Various methods of measuring similarity commonly known in the state of the art may be used, e.g. cosine similarity.

[0032] Fig. 3 shows an exemplary flow diagram of the image recognition model 330 classifying an input image 310. The image recognition model 330 may include a feature extraction module 320M and a classification module 340M. The feature extraction module 320M may include at least one feature extraction layer 322 (as described in Fig. 2). The feature extraction module 320 may be configured to extract a plurality of feature vectors of the feature items in the set of predetermined images.

[0033] Referring to Fig. 3, the classification module 340M may include a classification layer 342, which is a fully connected neural network layer. The classification layer 342 may be generated based on the predetermined set of images. For the generation of the classification layer 342, a plurality of categories may be determined for the image recognition model 330. For example, to classify images related to animals, the plurality of categories may include dogs, cats, sharks, etc. A plurality of feature vectors (the embeddings) may be extracted from the predetermined set of images by the feature extraction module 320M based on the feature items in the predetermined set of images. After that, the plurality of feature vectors may be grouped into the plurality of categories based on the categories of the corresponding predetermined set of images to form the feature vectors for each of the plurality of categories. At least one centroid 346 may be determined for each of the plurality of categories. A plurality of centroids 346 may be determined for each of the plurality of categories. One or more or a plurality of centroids 346 for each of the plurality of categories may be determined by clustering the feature vectors of the category into at least one cluster 322C and a centroid 346 for each of the at least one cluster may be generated. Consequently, a plurality of centroids 346 may be generated for the plurality of clusters 322C. The clustering can be performed using a clustering technique such as k-means, k-means++, k-medians, k-medoids, etc. As the number of centroids 346 increases, the accuracy of the recognition model may be improved. The classification layer 342 may be generated based on the centroids 346 of all the plurality of categories. Each centroid 346 may be represented by a feature vector, hence if there are a number of categories, C , and a number of centroids, K , or feature vectors, the classification layer 342, e.g. as a fully connected layer,

may be formed by $C \times K$ number of centroids 346 or feature vectors and be used to classify an input image 310 into one of the plurality of categories. The classification module 340M may be appended after the feature extraction module 320M to form the image recognition model 330. Specifically, the classification layer 342 may be appended to the image recognition model 330. The image recognition model 330 may be used to recognize the input image 310.

[0034] By clustering the plurality of feature vectors to the number of centroids, K , the classification module 340M may be configured to perform similarly to a search-based classification module where the feature database is compressed to reduce the size of the image recognition model 330. At the same time, by increasing the number of centroids, K , the accuracy of the classification module 340M is increased.

[0035] It is possible to vary the precision of the classification module 340M by varying the number of centroids, K , of each of the plurality of categories. The number of centroids, K , for each of the plurality of categories may be the same. The number of centroids may be the same for each of the plurality of classes in the classification module 340M. Varying the number of centroids, K , may also be known as adjusting the granularity of the classification layer 342. By adjusting the granularity of the classification layer 342, the accuracy and size of the classification layer 342 may be adjusted. Specifically, when the number of centroids, K , is increased, the accuracy of the classification layer 342 may be improved, while the size of the classification layer 342 is increased. Conversely, when the number of centroids, K , is reduced, the accuracy and size of the classification layer 342 may be decreased. Therefore, depending on the resource of the system where the image recognition model 330 is deployed, the granularity of the image recognition model 330 may be adjusted accordingly to achieve optimal results. For example, to vary the precision of the classification module, the number of clusters 322C of each of the plurality of categories may be varied into a new number of clusters 322C. When the number of clusters 322C are varied, the plurality of feature vectors of each of the plurality of categories may be re-clustered into the new number of clusters 322C. For example, the number of clusters 322C may be increased or decreased. Based on the new number of clusters 322C, at least one centroid for each of the new number of clusters 322C may be re-determined. As such, the number of centroids K may be varied. Noteworthy, after the number of centroids, K , is varied or adjusted, the plurality of feature vectors of each of the plurality of categories may be re-clustered into new clusters 322C accordingly. Based on the new number

of clusters 322C, new clustered centroids may be determined, and the classification layer 342 in the image recognition model 330 may be re-generated based on the new clustered centroids. All of these may be easily done without re-training the classification layers as well as the feature extraction layers.

[0036] Similarly, in the event that the image recognition model 330 needs to be updated, such as when a new category needs to be recognised, the classification layer 342 may be modified directly without re-training the whole image recognition model 330. To update the classification layer 342 to recognize the new category, a selected plurality of images that belong to the new category may be provided. A plurality of new feature vectors of the plurality of images may be extracted using the feature extraction module 320M. Next, the plurality of new feature vectors of the new category may be clustered into at least one new cluster 322C. For example, the plurality of new feature vectors may be clustered into a plurality of new clusters 322C. At least one centroid may be generated for each of the at least one new cluster 322C. Each of the at least one centroid may be represented by a feature vector and the corresponding feature vector of the at least one centroid or corresponding feature vectors of a plurality of the centroids may be appended to the classification layer 342, i.e. the classification layer 342 may be re-generated based on the plurality of feature vectors of the plurality of centroids of the plurality of categories, including the plurality of new feature vectors of the new category. Thereafter, the image recognition model 330 may be used to recognize images of the new category.

[0037] It may also be possible to update or improve the accuracy of one of the plurality of categories of the classification layer 342 by providing a new set of plurality of selected images of the one of the plurality of categories to the image recognition model 330. Plurality of selected images in one of the plurality of categories may be received by the system. Feature extraction module 320M may be configured to extract a plurality of new feature vectors of feature items from the plurality of selected images. The plurality of new feature vectors may be added to the existing set of feature vectors of the one of the plurality of categories to form a new set of feature vectors. Thereafter, the new set of feature vectors may be re-clustered into at least one cluster 322C. At least one centroid K may be re-generated for each of the at least one cluster 322C. Each of the at least one centroid K may be represented by a feature vector. The feature vector of each of the at least one centroid for each of the at least one cluster 322C may be

appended to the classification layer 342 of the image recognition model 330, i.e. the classification layer 342 may be re-generated based on the feature vector of the at least one cluster 322C and the plurality of feature vectors of the other plurality of categories. Thereafter, the one of the plurality of categories with the feature vectors of the new set of selected images may be updated or improved to recognise an input image in the one of the plurality of categories.

[0038] In summary, a new category may easily be added into the image recognition model 330, or an existing category can be fine-tuned by updating the corresponding centroids, all without ever re-training the at least one feature extraction layer 320. In this way, the image recognition model 330 is more flexible, controllable and easy to manage.

[0039] Fig. 4 shows a flow diagram of a method 2000 for classifying an image, the method 2000 includes receiving an input image in block 2100, inputting the input image to the generated image recognition model in block 2200 to classify the input image into one of the plurality of categories, and outputting a category name of the one of the plurality of categories of the input image from the image recognition model in block 2300.

[0040] Fig. 5 shows an exemplary flow diagram of the image recognition model 530 identifying an input image 510. As shown in Fig. 5, the input image 510 may be input to the image recognition model 530. The image recognition model 530 may, after recognising the input image 510, output the category name of the input image 510. When the input image 510 is input to the image recognition model 530, the feature extraction module 520 may extract the feature vector of the input image 510. Thereafter, the feature vector may be input to the classification module 540 where the classification layer 542 is configured to classify it into one of the plurality of categories C. To do so, the feature vector may be compared to the plurality of centroids in the classification layer 542. Classification layer 542 may be configured to determine one centroid that is nearest to the feature vector by calculating the similarity between the feature vector of the input image 510 and the plurality of centroids and return the centroid that is the most similar to the feature vector, e.g. by cosine distance. Accordingly, the category name of the returned centroid may be identified, and the name of the category, e.g. “shark”, may be output by the image recognition model 530.

[0041] The image recognition model 530 may be suitable for a system that has limited resources, e.g. a mobile device like a mobile phone. In addition, the image recognition model 530 is able to overcome the problems of complex deployment of a search-based recognition module as its structure is simpler and the parameters directly stored within the model itself. Also, the granularity of the classification layer 542 is more flexible, controllable and easily adjusted. Hence, the accuracy of the image recognition model 530 may be optimised for the resource available in the system. In fact, different models with different granularity (and hence accuracy) can easily be computed and deployed to different systems, if needed.

[0042] Fig. 6 shows a schematic view of an exemplary embodiment of the system for deploying the methods in Fig. 1. The abovementioned methods may be deployed in a system. System may include a mobile device, e.g. a mobile phone, tablet, laptop; a computer, etc. System includes a processor 610 and a memory 620 in communication to the processor 610 for storing instructions executable by the processor 610. System may further include at least one of a power source 630, a multimedia module 640, an audio module 650, an input/output (I/O) interface 660, a sensor module 670 and a communication module 680.

[0043] The processor 610 typically controls overall operations of the system 600, such as the operations associated with display, telephone calls, data communications, camera operations, and recording operations. The processor 610 may include one or more processors 1340 to execute instructions to perform all or part of the steps in the above described methods. Moreover, the processor 610 may include one or more modules which facilitate the interaction between the processor 610 and other modules. For instance, the processor 610 may include a multimedia module to facilitate the interaction between the multimedia module 640 and the processor 610. System may be in communication with a server 602 via a network.

[0044] Memory 620 may be configured to store various types of data to support the operation of the system 600. For example, the data may include instructions for any applications or above methods operated on the system 600, contact data, phonebook data, messages, pictures, video, etc. Memory 620 may be implemented using any type of volatile or non-volatile memory devices, or a combination thereof, such as a static random access memory (SRAM), an electrically erasable programmable read-only memory (EEPROM), an erasable programmable

read-only memory (EPROM), a programmable read-only memory (PROM), a read-only memory (ROM), a magnetic memory, a flash memory, a magnetic or optical disk.

[0045] Power source 630 provides power to various modules of the system 600. Power source 630 may include a power management system, one or more power sources, and any other modules associated with the generation, management, and distribution of power in the system 600.

[0046] Multimedia module 640 may include a screen providing an output interface between the system 600 and the user. In some embodiments, the screen may include a liquid crystal display (LCD), organic light-emitting diode (OLED), a touch panel, etc. If the screen includes the touch panel, the screen may be implemented as a touch screen to receive input signals from the user. Touch panel may include one or more touch sensors to sense touches, swipes, and gestures on the touch panel. The touch sensors may not only sense a boundary of a touch or swipe action, but also sense a period of time and a pressure associated with the touch or swipe action. In some embodiments, the multimedia module 640 may include a front camera and/or a rear camera for capturing the input image. The front camera and/or the rear camera may receive an external multimedia datum while the system 600 is in an operation mode, such as a photographing mode or a video mode. Each of the front camera and the rear camera may be a fixed optical lens system or have focus and optical zoom capability.

[0047] Audio module 650 may be configured to output and/or input audio signals. For example, the audio module 650 may include a microphone ("MIC") configured to receive an external audio signal when the system 600 is in an operation mode, such as a call mode, a recording mode, and a voice recognition mode. The received audio signal may be further stored in the memory 620 or transmitted via the communication module 680. In some embodiments, the audio module 650 further includes a speaker to output audio signals.

[0048] I/O interface 660 provides an interface between the processor 610 and peripheral interface modules, such as a keyboard, a click wheel, buttons, and the like. The buttons may include, but are not limited to, a home button, a volume button, a starting button, and a locking button.

[0049] Sensor module 670 includes one or more sensors to provide status assessments of various aspects of the system 600. For instance, the sensor module 670 may detect an open/closed status of the system 600, relative positioning of modules, e.g., the display and the keypad, of the system 600, a change in position of the system 600 or a module of the system 600, a presence or absence of user contact with the system 600, an orientation or an acceleration/deceleration of the system 600, and a change in temperature of the system 600, etc. Sensor module 670 may include a proximity sensor configured to detect the presence of nearby objects without any physical contact. Sensor module 670 may also include a light sensor, such as a CMOS or CCD image sensor, for use in imaging applications. In some embodiments, the sensor module 670 may also include an accelerometer sensor, a gyroscope sensor, a magnetic sensor, a pressure sensor, or a temperature sensor.

[0050] Communication module 680 may be configured to facilitate communication, wired or wirelessly, between the system 600 and other devices or the server 602. System 600 may access a wireless network based on a communication standard, such as WiFi, 2G, or 3G, LTE, and 4G cellular technologies or a combination thereof. In one exemplary embodiment, the communication module 680 may receive a broadcast signal or broadcast associated information from an external broadcast management system via a broadcast channel. In one exemplary embodiment, the communication module 680 may further include a near field communication (NFC) module to facilitate short-range communications. For example, the NFC module may be implemented based on a radio frequency identification (RFID) technology, an infrared data association (IrDA) technology, an ultra-wideband (UWB) technology, a Bluetooth (BT) technology, and other technologies.

[0051] In exemplary embodiments, the system 600 may be implemented with one or more application specific integrated circuits (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), controllers, micro-controllers, microprocessors, or other electronic modules, for performing the above described methods.

[0052] A non-transitory computer readable storage medium comprising instructions, wherein the instructions, when executed by a processor in a system, cause the system to append at least one feature extraction layer to an image recognition model, extract a plurality of feature vectors

from a set of pre-determined images, group the plurality of feature vectors into a plurality of categories, cluster the plurality of feature vectors of each of the plurality of categories into at least one cluster, determine at least one centroid for each of the at least one cluster, such that each of the at least one cluster comprises at least one centroid, such that each of the at least one centroid is represented by a feature vector, generate the classification layer based on the feature vector of the at least one centroid of the plurality of categories, and append the classification layer to the image recognition model.

[0053] A non-transitory computer readable storage medium comprising instructions, wherein the instructions, when executed by a processor in a system, cause the system to receive an input image, input the input image to the generated image recognition model as mentioned above to classify the input image into one of the plurality of categories, and output a category name of the one of the plurality of categories of the input image from the image recognition model.

[0054] A skilled person would appreciate that the features described in one example may not be restricted to that example and may be combined with any one of the other examples.

[0055] The present invention relates to a system and a method for generating an image recognition model and classifying an input image generally as herein described, with reference to and/or illustrated in the accompanying drawings.

Claim

1. A method of generating an image recognition model for recognising an input image, comprising:

appending at least one feature extraction layer to the image recognition model;

extracting a plurality of feature vectors from a set of predetermined images;

grouping the plurality of feature vectors into a plurality of categories;

clustering the plurality of feature vectors of each of the plurality of categories into at least one cluster;

determining at least one centroid for each of the at least one cluster, wherein each of the at least one cluster comprises at least one centroid, wherein each of the at least one centroid is represented by a feature vector;

generating a classification layer based on the feature vector of the at least one centroid of the plurality of categories; and

appending the classification layer to the image recognition model.

2. The method according to claim 1, comprising varying the number of clusters of each of the plurality of categories into a new number of clusters, re-clustering the plurality of feature vectors of each of the plurality of categories into the new number of clusters, re-determining at least one centroid for each of the new number of clusters wherein each of the at least one centroid is represented by a feature vector, and re-generating the classification layer of the image recognition model based on the feature vector of the at least one centroid of each of the new number of clusters.

3. The method according to claim 1 or 2, further comprising receiving a plurality of images for a new category, extracting a plurality of new feature vectors from the plurality of images by the feature extraction module, clustering the plurality of new feature vectors into at least one new cluster, generating at least one centroid for each of the at least one new cluster, wherein each of the at least one centroid is represented by a feature vector, and appending the feature vector of each of the at least one centroid for each of the at least one new cluster to the classification layer of the image recognition model.

4. The method according to any one of claim 1 to 3, further comprising receiving a plurality of selected images in one of the plurality of categories, extracting a plurality of new feature vectors of the plurality of selected images, adding the plurality of new feature vectors to the plurality of feature vectors of the one of the plurality of categories to form a new set of feature vectors, re-clustering the new set of feature vectors into at least one cluster, re-generating at least one centroid for each of the at least one cluster, wherein each of the at least one centroid is represented by a feature vector, and appending the feature vector of each of the at least one centroid for each of the at least one cluster to the classification layer of the image recognition model.
5. The method according to any one of claims 1 to 4, wherein the plurality of feature vectors are clustered using the k-mean algorithm.
6. A method for classifying an input image, the method comprising:
 - receiving the input image;
 - inputting the image to the generated image recognition model as claimed in any one of claims 1 to 5 to classify the input image into one of the plurality of categories;
 - and
 - outputting a category name of the one of the plurality of categories of the input image from the image recognition model.
7. The method according to claim 6, wherein the method comprises:
 - extracting a feature vector of the input image;
 - comparing the feature vector to the plurality of centroids in the classification layer;
 - determining one of the plurality of centroids that is nearest to the feature vector;
 - identifying the category of the one of the plurality of centroids; and
 - retrieving a category name of the category of the one of the plurality of centroids.
8. The method according to claim 7, wherein comparing the feature vector to the plurality of centroids comprises calculating the distances between the feature vector and each of the feature vector of the at least one centroid of the plurality of categories.

9. A system for generating an image recognition model for recognising an input image, the system comprising:

a processor; and

a memory in communication to the processor for storing instructions executable by the processor,

wherein the processor is configured to:

append at least one feature extraction layer to an image recognition model;

extract a plurality of feature vectors from a set of pre-determined images;

group the plurality of feature vectors into a plurality of categories;

cluster the plurality of feature vectors of each of the plurality of categories into at least one cluster;

determine at least one centroid for each of the at least one cluster, wherein each of the at least one cluster comprises at least one centroid, wherein each of the at least one centroid is represented by a feature vector;

generate the classification layer based on the feature vector of the at least one centroid of the plurality of categories; and

append the classification layer to the image recognition model.

10. The system according to claim 9, wherein the processor is configured to vary the number of clusters of each of the plurality of categories into a new number of clusters, re-cluster the plurality of feature vectors of each of the plurality of categories into the new number of clusters, re-determine at least one centroid for each of the new number of clusters wherein each of the at least one centroid is represented by a feature vector, re-generate the classification layer in the recognition model based on the feature vector of the at least one centroid of each of the new number of clusters.

11. The system according to claim 9 or 10, wherein the processor is configured to receive a plurality of selected images in one of the plurality of categories, extract a plurality of new

feature vectors of the plurality of selected images, add the plurality of new feature vectors to the plurality of feature vectors of one of the plurality of categories to form a new set of feature vectors, re-cluster the new set of feature vectors into at least one cluster, re-generate at least one centroid for each of the at least one cluster, wherein each of the at least one centroid is represented by a feature vector, and append the feature vector of each of the at least one centroid for each of the at least one cluster to the classification layer to the image recognition model.

12. The system according to any one of claims 9 to 11, wherein the plurality of feature vectors are clustered using k-mean algorithm.

13. A system for classifying an input image, the system comprising:

a processor; and

a memory in communication to the processor for storing instructions executable by the processor,

wherein the processor is configured to:

receive the input image;

input the input image to the generated image recognition model as claimed in any one of claims 9 to 11 to classify the input image into one of the plurality of categories; and

output a category name of the one of the plurality of categories of the input image from the image recognition model.

14. The system according to claim 13, wherein the processor is configured to extract a feature vector of the input image, compare the feature vector to the plurality of centroids in the classification layer, determine one of the plurality of centroids that is nearest to the feature vector, identify the category of the one of the plurality of centroids and retrieve the category name of the category of the one of the plurality of centroids.

15. A non-transitory computer readable storage medium comprising instructions, wherein the instructions, when executed by a processor in a system, cause the system to:

append at least one feature extraction layer to an image recognition model;

extract a plurality of feature vectors from a set of pre-determined images;

group the plurality of feature vectors into a plurality of categories;
cluster the plurality of feature vectors of each of the plurality of categories into at least one cluster;
determine at least one centroid for each of the at least one cluster, wherein each of the at least one cluster comprises at least one centroid, wherein each of the at least one centroid is represented by a feature vector;
generate the classification layer based on the feature vector of the at least one centroid of the plurality of categories; and
append the classification layer to the image recognition model.

16. A non-transitory computer readable storage medium comprising instructions, wherein the instructions, when executed by a processor in a system, cause the system to:

receive an input image;
input the input image to the generated image recognition model as claimed in any one of claims 9 to 11 to classify the input image into one of the plurality of categories; and
output a category name of the one of the plurality of categories of the input image from the image recognition model.

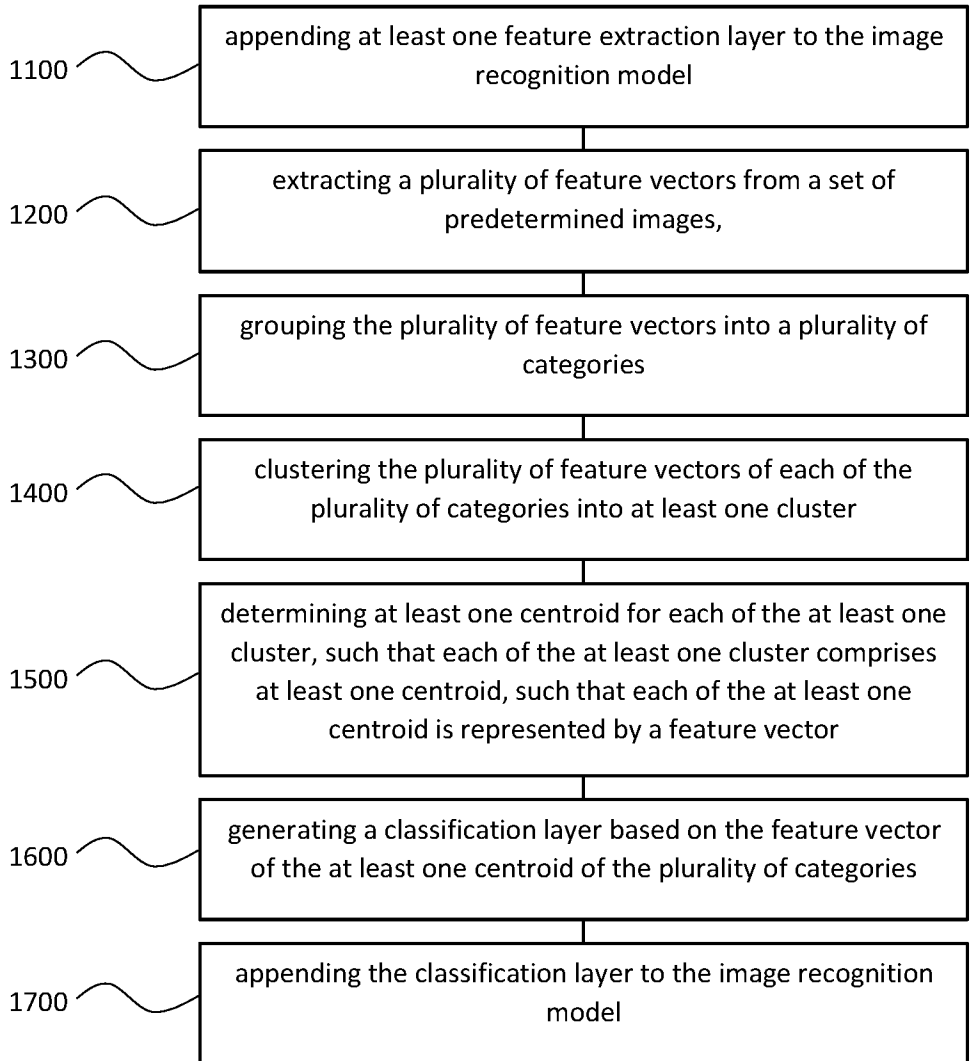


Fig. 1

230

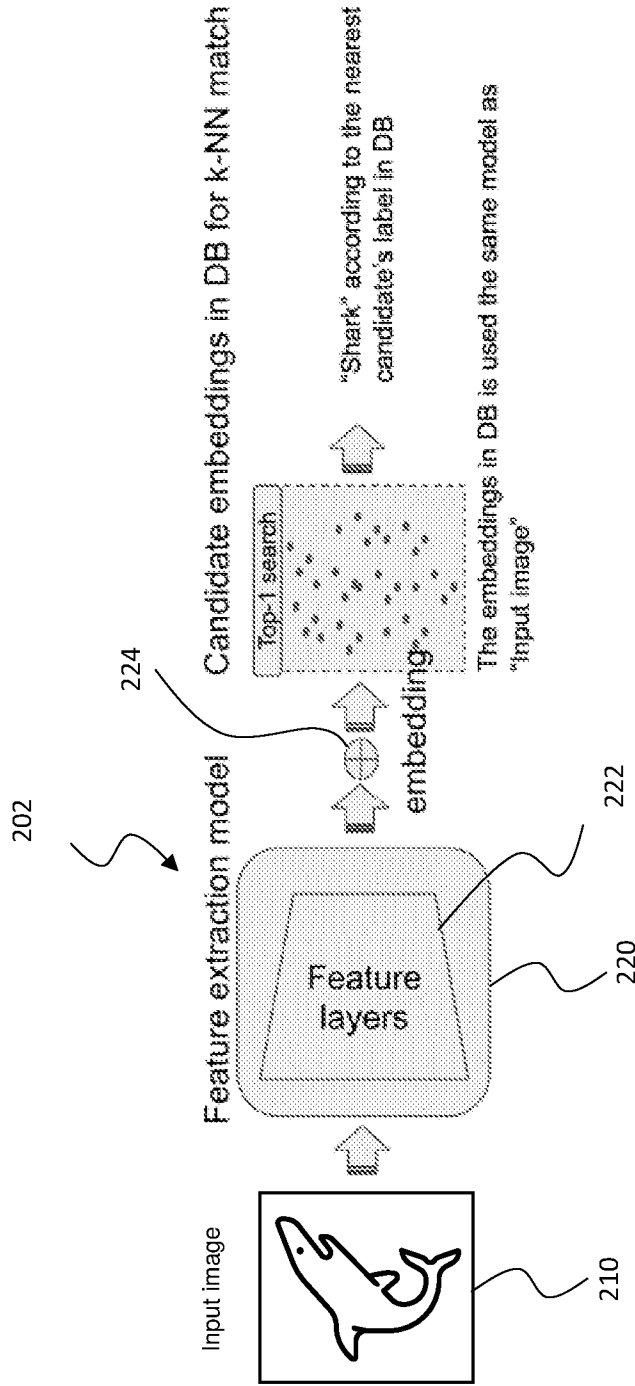


Fig. 2

330

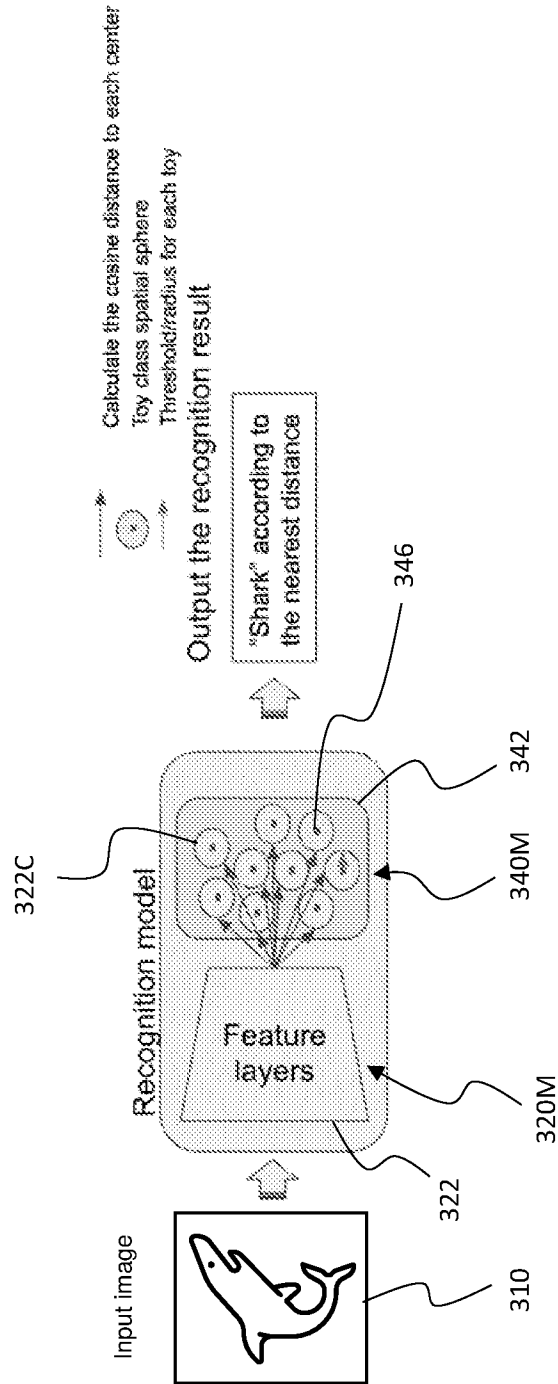


Fig. 3

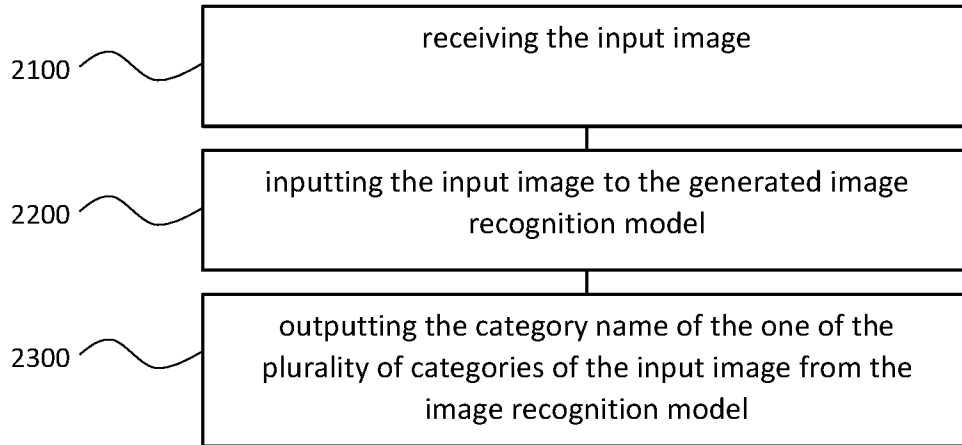


Fig. 4

530

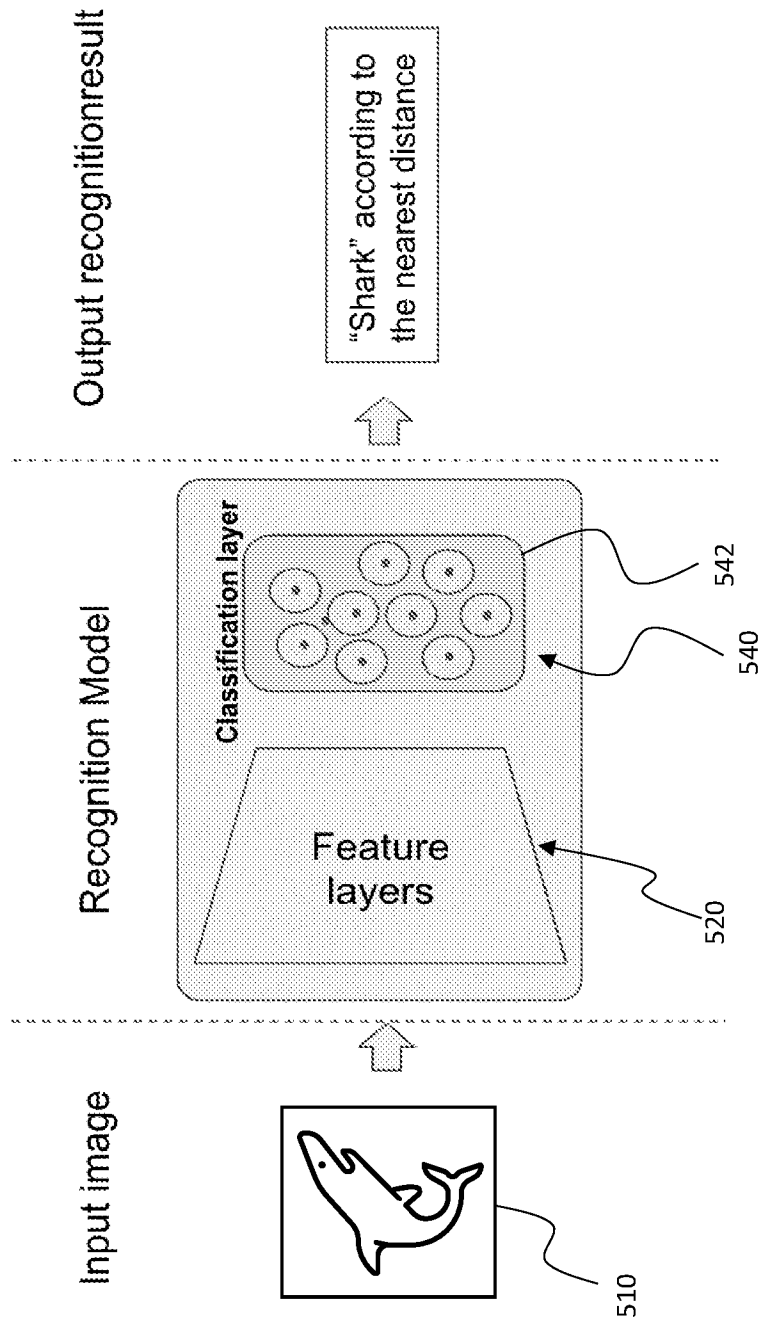


Fig. 5

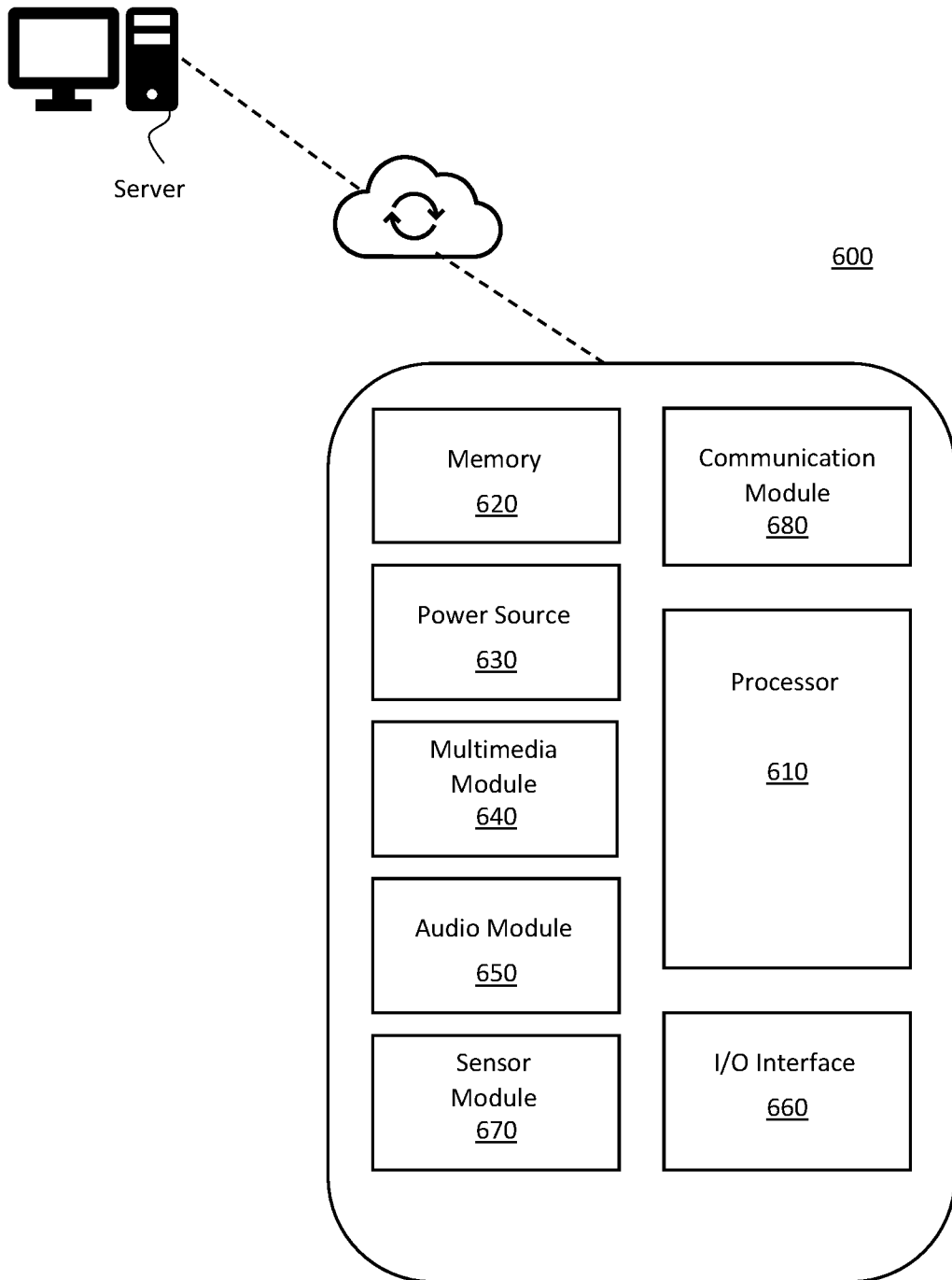


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2020/050642

A. CLASSIFICATION OF SUBJECT MATTER

G06N 20/00 (2019.01) G06N 3/02 (2006.01) G06K 9/62 (2006.01) G06T 7/33 (2017.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Google, Google Scholar, Google Patents, www.freepatentsonline.com, EPOQUE (PATENW) with IPC/CPC marks G06K9/6272: image, recognition, model, centroid, classification, extraction, layer, append, feature, vector, centroid and other similar terms

Applicant(s)/Inventor(s) name searched in internal databases provided by IP Australia

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

 Further documents are listed in the continuation of Box C See patent family annex

* Special categories of cited documents:		
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	
"D" document cited by the applicant in the international application	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family	
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

21 January 2021

Date of mailing of the international search report

21 January 2021

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
Email address: pct@ipaustralia.gov.au

Authorised officer

Andrew Munro
AUSTRALIAN PATENT OFFICE
(ISO 9001 Quality Certified Service)
Telephone No. +61262850715

INTERNATIONAL SEARCH REPORT

International application No.

C (Continuation).

DOCUMENTS CONSIDERED TO BE RELEVANT

PCT/SG2020/050642

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2012/0170852 A1 (ZHANG et al) 05 July 2012 see, for example, the abstract, paragraphs [0030] and [0032] and claim 1	1-16
A	BAKER, B., et. al., 'Designing Neural Network Architectures using Reinforcement Learning' [downloaded from the Internet 6 January 2021] <URL: https://arxiv.org/abs/1611.02167 > published 2017 see, for example, the abstract, section 7 (second paragraph), section 4 and figure 2	1-16
A	US 2017/0262737 A1 (MAGIC LEAP, INC.) 14 September 2017 see, for example, the abstract, paragraph [0014] and claim 1	1-16
A	US 9996755 B2 (TSAI et al) 12 June 2018 see, for example, the abstract and claim 1	1-16
A	US 5796924 A (ERRICO et al) 18 August 1998 see, for example, the abstract, column 1 lines 25-37, column 4 lines 1-9 and claims 1 and 11	1-16

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2020/050642

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	Publication Date	Publication Number	Publication Date
US 2012/0170852 A1	05 July 2012	US 2012170852 A1	05 Jul 2012
		TW 201112134 A	01 Apr 2011
		TW I484423 B	11 May 2015
		WO 2011037579 A1	31 Mar 2011
US 2017/0262737 A1	14 September 2017	US 2017262737 A1	14 Sep 2017
		US 10255529 B2	09 Apr 2019
		AU 2017230184 A1	06 Sep 2018
		CA 3015658 A1	14 Sep 2017
		CN 108780519 A	09 Nov 2018
		EP 3427192 A1	16 Jan 2019
		JP 2019511777 A	25 Apr 2019
		JP 2020091922 A	11 Jun 2020
		KR 20180117704 A	29 Oct 2018
		KR 20200035499 A	03 Apr 2020
		US 2019286951 A1	19 Sep 2019
		WO 2017156547 A1	14 Sep 2017
		US 9996755 B2	12 June 2018
US 9996755 B2	12 Jun 2018		
TW I590197 B	01 Jul 2017		
TW 201804437 A	01 Feb 2018		
US 5796924 A	18 August 1998	US 5796924 A	18 Aug 1998
		AU 2209397 A	10 Oct 1997
		FR 2746530 A1	26 Sep 1997
		TW 343315 B	21 Oct 1998
		WO 9735264 A1	25 Sep 1997

End of Annex

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

Form PCT/ISA/210 (Family Annex)(July 2019)