(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 December 2017 (21.12.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/214968 A1

(51) International Patent Classification:
G06K 9/62 (2006.01)

(21) International Application Number:
PCT/CN2016/086152

(22) International Filing Date:
17 June 2016 (17.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **NOKIA TECHNOLOGIES OY** [FI/FI]; Karaportti 3, 02610 Espoo (FI).

(71) Applicant *(for LC only)*: **NOKIA TECHNOLOGIES (BEIJING) CO., LTD.** [CN/CN]; Unit 18, Room 1001, Level 10, East Tower, World Financial Center, No.1, East Third, Ring Middle Road, Chaoyang District, Beijing 100020 (CN).

(72) Inventor: **CAO, Jiale**; Room 6-4-501, Xin-Yuan-Cun Village, Nankai District, Tianjin 300072 (CN).

(74) Agent: **KING & WOOD MALLESONS**; 20th Floor, East Tower, World Financial Centre, No.1 Dongsanhuan Zhonglu, Chaoyang District, Beijing 100020 (CN).

(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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*

(54) Title: METHOD AND APPARATUS FOR CONVOLUTIONAL NEURAL NETWORKS

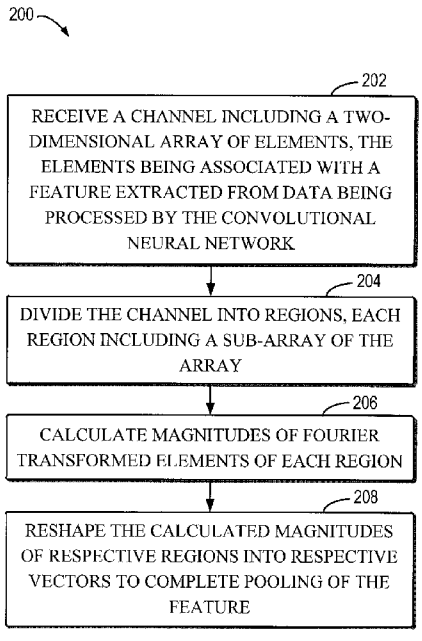

Fig. 2

(57) Abstract: Embodiments of the present disclosure relate to method and apparatus for convolutional neural networks. In example embodiments, a method implemented at a pooling layer of a convolutional neural network comprises receiving a channel including a two-dimensional array of elements, the elements being associated with a feature extracted from data being processed by the convolutional neural network. The method further comprises dividing the channel into regions, each region including a sub-array of the array. The method further comprises calculating magnitudes of Fourier transformed elements of each region. The method further comprises reshaping the calculated magnitudes of respective regions into respective vectors to complete pooling of the feature.

WO 2017/214968 A1

METHOD AND APPARATUS FOR CONVOLUTIONAL NEURAL NETWORKS

FIELD

[0001] Embodiments of the present disclosure generally relate to the field of convolutional neural networks, and in particular, to a method and apparatus for a pooling layer of a convolutional neural network.

BACKGROUND

[0002] Convolutional neural networks (CNNs) have achieved state-of-the-art performance in the applications of image recognition, object detection, acoustic recognition, and so on. Representative applications of convolutional Neural Networks include, but are not limited to, AlphaGo, Advanced Driver Assistance Systems (ADAS), self-driving cars, Optical Character Recognition (OCR), face recognition, large-scale image classification (for example, ImageNet classification), and Human Machine Interaction (HCI). Convolutional neural networks are mainly organized in interweaved layers of two types: convolutional layers and pooling (subsampling) layers, with a convolutional layer (or several convolutional layers) followed by a pooling layer.

[0003] Pooling is a process that replaces the output of its corresponding convolutional layers at certain location with summary statistic of the nearby outputs. Pooling over spatial regions contributes to make feature representation become shift invariant and also contributes to improve the computational efficiency of the convolutional neural network.

[0004] Currently, max-pooling and average pooling (also called sum pooling) are two dominant pooling methods. The max-pooling operation outputs the maximum value within a rectangular neighborhood. The average pooling operation outputs the average value within a rectangular neighborhood. In some scenarios, the max-pooling and the average pooling can be combined.

[0005] In addition, stochastic pooling is proposed in which the deterministic pooling operations are replaced with a stochastic procedure by randomly picking the activation within each pooling region according to a multinomial distribution. In another traditional pooling method, the pooling formulates a kind of convolution. Moreover, others

proposed a spatial pyramid pooling to deal with a situation where the training images are of different size.

SUMMARY

[0006] In general, example embodiments of the present disclosure provide a method and apparatus for a pooling layer of a convolutional neural network.

[0007] In a first aspect, a method implemented at a pooling layer of a convolutional neural network is provided. The method comprises receiving a channel including a two-dimensional array of elements, the elements being associated with a feature extracted from data being processed by the convolutional neural network. The method further comprises dividing the channel into regions, each region including a sub-array of the array. The method further comprises calculating magnitudes of Fourier transformed elements of each region. The method further comprises reshaping the calculated magnitudes of respective regions into respective vectors to complete pooling of the feature.

[0008] In some embodiments, the receiving may comprise receiving the channel from a convolutional layer of the convolutional neural network.

[0009] In some embodiments, the dividing may comprise dividing the channel into non-overlapping regions.

[0010] In some embodiments, each region may have a same quantity of elements.

[0011] In some embodiments, each region may have n×n elements, where n is an integer equal to or greater than two.

[0012] In some embodiments, the reshaping comprises arranging the calculated magnitudes along the channel dimension.

[0013] In some embodiments, the arranging may comprise arranging the calculated magnitudes of the respective regions in a same order.

[0014] In some embodiments, the data may be associated with at least one of image recognition, acoustic recognition, object detection, Advanced Driver Assistance Systems (ADAS), self-driving cars, Optical Character Recognition (OCR), face recognition, large-scale image classification, and Human Machine Interaction (HCI).

[0015] In a second aspect, an apparatus implemented at a pooling layer of a convolutional neural network is provided. The apparatus comprises at least one processor and at least one memory including computer program code. The at least one memory and the computer program code are configured to, with the at least one processor, cause the apparatus at least to perform: receiving a channel including a two-dimensional array of elements, the elements being associated with a feature extracted from data being processed by the convolutional neural network; dividing the channel into regions, each region including a sub-array of the array; calculating magnitudes of Fourier transformed elements of each region; and reshaping the calculated magnitudes of respective regions into respective vectors to complete pooling of the feature.

[0016] In some embodiments, the at least one memory and the computer program code may further be configured to, with the at least one processor, cause the apparatus to perform receiving the channel from a convolutional layer of the convolutional neural network.

[0017] In some embodiments, the at least one memory and the computer program code may further be configured to, with the at least one processor, cause the apparatus to perform dividing the channel into non-overlapping regions.

[0018] In some embodiments, each region may have a same quantity of elements.

[0019] In some embodiments, each region may have n×n elements, where n is an integer equal to or greater than two.

[0020] In some embodiments, the at least one memory and the computer program code may further be configured to, with the at least one processor, cause the apparatus to perform: reshaping the calculated magnitudes of respective regions by arranging the calculated magnitudes along the channel dimension.

[0021] In some embodiments, the at least one memory and the computer program code may further be configured to, with the at least one processor, cause the apparatus to perform: arranging the calculated magnitudes of the respective regions in a same order.

[0022] In some embodiments, the data may be associated with at least one of image recognition, acoustic recognition, object detection, Advanced Driver Assistance Systems (ADAS), self-driving cars, Optical Character Recognition (OCR), face recognition,

large-scale image classification, and Human Machine Interaction (HCI).

[0023] In a third aspect, an apparatus is provided. The apparatus comprises means for performing a method according to the first aspect.

[0024] In a fourth aspect, a computer program product is provided. The computer program product comprises at least one computer readable non-transitory memory medium having program code stored thereon. The program code, when executed by an apparatus, causes the apparatus to perform a method according to the first aspect.

[0025] Compared to existing pooling methods, the proposed pooling method and apparatus can achieve complete shift-invariance. Moreover, instead of representing a region with a single value, the proposed pooling method and apparatus represent a region with a vector, and thus they do not lose useful information.

[0026] It is to be understood that the summary section is not intended to identify key or essential features of embodiments of the present disclosure, nor is it intended to be used to limit the scope of the present disclosure. Other features of the present disclosure will become easily comprehensible through the following description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] Through the more detailed description of some embodiments of the present disclosure in the accompanying drawings, the above and other objects, features and advantages of the present disclosure will become more apparent, wherein:

[0028] Fig. 1 is a schematic diagram of an environment in which embodiments of the present disclosure can be implemented;

[0029] Fig. 2 is a flowchart illustrating a method implemented at a pooling layer of a convolutional neutral network in accordance with embodiments of the present disclosure;

[0030] Fig. 3 is a schematic diagram illustrating a method implemented at a pooling layer of a convolutional neutral network in accordance with embodiments of the present disclosure;

[0031] Fig. 4 is a block diagram illustrating an apparatus implemented at a pooling layer of a convolutional neutral network in accordance with an embodiment of the present disclosure; and

[0032] Fig. 5 is a block diagram illustrating an apparatus implemented at a pooling layer of a convolutional neutral network in accordance with another embodiment of the present disclosure.

[0033] Throughout the drawings, the same or similar reference numerals represent the same or similar elements.

DETAILED DESCRIPTION

[0034] Hereinafter, the principle and spirit of the present disclosure will be described with reference to illustrative embodiments. It should be understood, all these embodiments are given merely for one skilled in the art to better understand and further practice the present disclosure, but not for limiting the scope of the present disclosure. For example, features illustrated or described as part of one embodiment may be used with another embodiment to yield still a further embodiment. In the interest of clarity, not all features of an actual implementation are described in this specification.

[0035] References in the specification to "one embodiment," "an embodiment," "an example embodiment," etc. indicate that the embodiment described may include a particular feature, structure, or characteristic, but it is not necessary that every embodiment includes the particular feature, structure, or characteristic. Moreover, such phrases are not necessarily referring to the same embodiment. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it is submitted that it is within the knowledge of one skilled in the art to affect such feature, structure, or characteristic in connection with other embodiments whether or not explicitly described.

[0036] It shall be understood that, although the terms "first" and "second" etc. may be used herein to describe various elements, these elements should not be limited by these terms. These terms are only used to distinguish one element from another. For example, a first element could be termed a second element, and similarly, a second element could be termed a first element, without departing from the scope of example embodiments. As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed terms.

[0037] The terminology used herein is for the purpose of describing particular

embodiments only and is not intended to be liming of example embodiments. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises", "comprising", "has", "having", "includes" and/or "including", when used herein, specify the presence of stated features, elements, and/or components etc., but do not preclude the presence or addition of one or more other features, elements, components and/ or combinations thereof.

[0038] In the following description and claims, unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skills in the art to which this disclosure belongs.

[0039] The example embodiments of the present disclosure will now be described with reference to Figs. 1-5. It is to be understood that these embodiments are described for the purpose of illustration only and help those skilled in the art to understand and implement the present disclosure, without suggesting any limitations as to the scope of the disclosure. The disclosure described herein can be implemented in various manners other than those described below.

[0040] Fig. 1 shows a schematic diagram of an environment in which embodiments of the present disclosure can be implemented. As shown in Fig. 1, an example framework of a convolutional neural network 100 may include an input color image 110, a first convolutional layer 120, a first pooling layer 130, a second convolutional layer 140, a second pooling layer 150, and a classification unit 160.

[0041] Although an image recognition scenario is depicted as an example application of the convolutional neural network 100, it is to be understood that the convolutional neural network 100 can also apply to any other suitable environments as described above. For example, the convolutional neural network 100 may be equivalently or similarly applied to acoustic recognition, object detection, Advanced Driver Assistance Systems (ADAS), self-driving cars, Optical Character Recognition (OCR), face recognition, large-scale image classification, Human Machine Interaction (HCI), and so on.

[0042] Further, although the convolutional neural network 100 is depicted as comprising two convolutional layers and two pooling layers, the convolutional neural network 100 may have more or less convolutional layers and pooling layers. Additionally, the

convolutional neural network 100 may include other layers or units which are not shown in Fig. 1 for simplicity.

[0043] In the image recognition process as shown in Fig. 1, after the input color image 110 is inputted to the convolutional neural network 100, a first convolution operation 115 may be performed on the input color image 110 and the first convolutional layer 120 may be generated accordingly. Then, a first pooling operation 125 may be performed on the first convolutional layer 120, and the first pooling layer 130 may thus be generated. In a same manner, a second convolution operation 135 may be performed on the first pooling layer 130 to generate the second convolutional layer 140. A second pooling operation 145 may be performed on the second convolutional layer 140 to generate the second pooling layer 150. As a result, the second pooling layer 150 may be classified by the classification unit 160 and the image recognition process may be accomplished.

[0044] Below are two roles of the pooling operations performed at a first pooling layer 130 and a second pooling layer 150. First, it makes the feature representation become invariant to small shifts of an input. This means that even if the input is shifted by a small amount, the values of most of the pooled outputs do not change. Second, by utilizing a single value to represent a region consisting of several elements, the pooling operation is able to subsample the input layers, for example,, to reduce the spatial size of the input layers.

[0045] Example embodiments of the present disclosure are aimed at presenting a complete shift-invariant pooling algorithm. Further, in the method and apparatus in accordance with example embodiments of the present disclosure, not only full shift-invariance property can be obtained but also no information is discarded. In particular, the first pooling layer 130 and/or the second pooling layer 150 may employ the pooling method as proposed herein to improve the performance of the convolutional neural network 100. This will be discussed in the following paragraphs. First of all, the pooling method as proposed herein will be described in detail with reference to Fig. 2.

[0046] Fig. 2 shows a flowchart illustrating a method 200 implemented at a pooling layer of a convolutional neutral network in accordance with embodiments of the present disclosure. The method 200 can be implemented by the first pooling layer 130 and/or the second pooling layer 150 as shown in Fig. 1, for example.

[0047] In step 202, a channel including a two-dimensional array of elements is received, the elements being associated with a feature extracted from data being processed by the convolutional neural network. In some embodiments, the channel may be a feature map of a convolutional layer associated with the pooling layer at which the method 200 is implemented. The feature may be extracted from the data by the convolutional layer. In these embodiments, step 202 may further comprise receiving the channel from a convolutional layer of the convolutional neural network.

[0048] In some embodiments, the data may be associated with at least one of image recognition, acoustic recognition, object detection, Advanced Driver Assistance Systems (ADAS), self-driving cars, Optical Character Recognition (OCR), face recognition, large-scale image classification, and Human Machine Interaction (HCI). This may depend on the specific application of the convolutional neural network.

[0049] In step 204, the channel is divided into regions, each region including a sub-array of the array. The pooling operation will be based on these regions. In other words, the pooling operation will be performed region by region, and the elements in a region will be pooled in the pooling of the region. The step 204 may further comprise dividing the channel into non-overlapping regions. In this manner, the performance of the pooling operation may be improved, and the complexity of the pooling operation may be reduced. In some embodiments, each region may have a same quantity of elements, such that the performance and the complexity of the pooling operation may further be improved. In other embodiments, each region may have n×n elements, where n is an integer equal to or greater than two.

[0050] In step 206, magnitudes of Fourier transformed elements of each region are calculated. The Fourier transform may be performed on elements of each region, and the same quantity of Fourier transformed elements may thus be generated. Each Fourier transformed element may have a magnitude and a phase. In some embodiments, both the magnitudes and phases of Fourier transformed elements of each region may be calculated, and then the phases may be dropped in step 206.

[0051] In step 208, the calculated magnitudes of respective regions are reshaped into respective vectors to complete pooling of the feature. It is to be understood that the quantity of the calculated magnitudes of a region is equal to that of the element in the

region. Thus, if there are a number of elements in a region, there are the same number of calculated magnitudes in the region. This number of calculated magnitudes may be reshaped into a vector, and the dimensionality of the vector is the quantity of the calculated magnitudes.

[0052] In some embodiments, the step 208 may further comprises arranging the calculated magnitudes along the channel dimension, so that different calculated magnitudes of a region belong to different channels. In these embodiments, the step 208 may further comprise arranging the calculated magnitudes of the respective regions in a same order. For example, if two calculated magnitudes belong to two different regions but are in a same channel, their related elements are located in the same position (such as, the same row and the same column) in the two different regions.

[0053] It can be seen that the pooling method proposed herein achieves complete shift-invariance due to the inherent shift-invariant nature of the Fourier transform. Moreover, instead of representing a region with a single value, the proposed pooling method represent a region with a vector. Therefore, the proposed pooling method does not lose useful information.

[0054] Fig. 3 shows a schematic diagram 300 illustrating a method implemented at a pooling layer of a convolutional neutral network in accordance with embodiments of the present disclosure. As shown in Fig. 3, a channel (for example, a feature map) 310 of a convolutional layer of a convolutional neutral network may be inputted to a pooling layer of the convolutional neutral network. As depicted, the channel 310 may have a two-dimensional array of elements 311. For example, the two-dimensional array in Fig. 3 is a 9×6 array, namely there are 9 rows and 6 columns of elements 311 in the channel 310. This is merely an example implementation without suggesting any limitations as to the scope of the present disclosure.

[0055] In a dividing step which is depicted by the dotted lines, the elements 311 of the channel 310 may be divided into regions 312, and the elements 311 of the channel 310 are pooled region by region. As an example, each region 312 may include 3×3 elements in Fig. 3, and the regions 312 may not be overlapped. Therefore, the channel 310 may be divided into 6 regions 312. Also, this is merely an example implementation without suggesting any limitations as to the scope of the present disclosure.

[0056] In step 315, the Fourier transform may be applied to each region 312, in order to calculate the magnitudes of the Fourier transformed elements of each region 312. In step 320, the 3×3 magnitudes of the result of the Fourier transform may be calculated. In step 325, the 3×3 magnitudes may be reshaped into a 9-dimensional vector (3×3 = 9). The quantity of 3×3 is merely an example implementation without suggesting any limitations as to the scope of the present disclosure.

[0057] As shown in Fig. 3, after the reshaping step 325, the elements 311 in each region 312 may be reshaped into respective 9-dimensional vectors 313 along the channel dimension. At last, the pooled result may be 9 new channels 330, and each channel 330 may include 6 respective magnitudes of 6 respective regions 312. The particular quantities here are merely an example implementation without suggesting any limitations as to the scope of the present disclosure.

[0058] Referring back to Fig. 1, as described above, the first pooling layer 130 and/or the second pooling layer 150 as shown in Fig. 1 may employ the pooling method as proposed herein to improve the performance of the convolutional neural network 100. Similarly with traditional convolutional neutral networks, a proposed Fourier Transform based pooling operation (FT-pooling operation) may be followed by a convolution operation.

[0059] As shown in Fig. 1, an H×W×3 color image 110 may be inputted to the convolutional neutral network 100. In other words, the color image 110 may have H rows, W columns, and 3 channels. The color image 110 may be convolved with a $a_1 \times b_1 \times 3$ filter in a first convolutional operation 115 and the resulting first convolutional layer 120 may contain k_1 channels (feature maps), with each channel being of size of H×W. The first convolutional layer 120 may be processed by a first FT-pooling operation 125. Because the pooling size may be n×n, the resulting first pooling layer 130 may have $k_2 = n \times k_1$ channels and each channel is of size (H/n)×(W/n).

[0060] Then, the first pooling layer 130 may be convolved with a $a_2 \times b_2 \times k_2$ filter in a second convolutional operation 135 and the resulting second convolutional layer 140 may contain k_3 channels (feature maps), with each channel being of size of (H/n)×(W/n). Subsequently, the second convolutional layer 140 may be processed by a second FT-pooling operation 145. Because the pooling size is n×n, the resulting second pooling layer 150 may have $k_4 = n \times k_3$ channels and each channel is of size $(H/n^2) \times (W/n^2)$. This

process may continue until a predefined number of convolutional and pooling layers is obtained. Finally, a classification 160 may be conducted at the last layer.

[0061] By employing the Fourier Transform based pooling method in accordance with example embodiments of the present disclosure, the performance of the convolutional neutral network 100 can be greatly improved, because the Fourier Transform based pooling method is completely shift-invariant and does not lose useful information.

[0062] Fig. 4 shows a block diagram illustrating an apparatus 400 implemented at a pooling layer of a convolutional neutral network in accordance with an embodiment of the present disclosure. The apparatus 400 may be operable to implement the example method 200 described with reference to Fig. 2 and possibly any other processes or methods. It is to be understood that the method 200 is not necessarily implemented by the apparatus 400. At least some steps of the method 200 may be performed by one or more other entities.

[0063] Particularly, as illustrated in Fig. 4, the apparatus 400 includes a receiving unit 401, a dividing unit 402, a calculating unit 403, and a reshaping unit 404. The receiving unit 401 is configured to receive a channel including a two-dimensional array of elements, the elements being associated with a feature extracted from data being processed by the convolutional neural network. The dividing unit 402 is configured to divide the channel into regions, each region including a sub-array of the array. The calculating unit 403 is configured to calculate magnitudes of Fourier transformed elements of each region. The reshaping unit 404 is configured to reshape the calculated magnitudes of respective regions into respective vectors to complete pooling of the feature.

[0064] In some embodiments, the receiving unit 401 may further be configured to receive the channel from a convolutional layer of the convolutional neural network. The dividing unit 402 may further be configured to divide the channel into non-overlapping regions. In some embodiments, each region may have a same quantity of elements. In some other embodiments, each region may have n×n elements, where n is an integer equal to or greater than two.

[0065] In some embodiments, the reshaping unit 404 may further be configured to arrange the calculated magnitudes along the channel dimension. In some other embodiments, the reshaping unit 404 may further be configured to arrange the calculated

magnitudes of the respective regions in a same order.

[0066] In some embodiments, the data is associated with at least one of image recognition, acoustic recognition, object detection, Advanced Driver Assistance Systems (ADAS), self-driving cars, Optical Character Recognition (OCR), face recognition, large-scale image classification, and Human Machine Interaction (HCI).

[0067] Fig. 5 shows a block diagram illustrating an apparatus 500 implemented at a pooling layer of a convolutional neutral network in accordance with another embodiment of the present disclosure. The apparatus 500 may be operable to implement the example method 200 described with reference to Fig. 2 and possibly any other processes or methods. It is to be understood that the method 200 is not necessarily implemented by the apparatus 500. At least some steps of the method 200 may be performed by one or more other entities.

[0068] The apparatus 500 may include at least one processor 501, such as a data processor (DP), and at least one memory 502 coupled to the processor 501. The memory 502 may be non-transitory machine readable storage medium and it may store a program (PROG) 503. The PROG 503 may include instructions that, when executed on the associated processor 501, enable the apparatus 500 to operate in accordance with the embodiments of the present disclosure, for example to perform the method 200. The embodiments herein may be implemented by computer software executable by the processor 501 of the apparatus 500, or by hardware, or by a combination of software and hardware. A combination of the at least one processor 501 and the at least one memory 502 may form processing means 510 adapted to implement various embodiments of the present disclosure, for example, method 200.

[0069] The memory 502 may be of any type suitable to the local technical environment and may be implemented using any suitable data storage technology, such as semiconductor based memory devices, magnetic memory devices and systems, optical memory devices and systems, fixed memory and removable memory, as non-limiting examples. While only one memory 502 is shown in the apparatus 500, there may be several physically distinct memory modules in the apparatus 500. The processor 501 may be of any type suitable to the local technical environment, and may include one or more of general purpose computers, special purpose computers, microprocessors, digital

signal processors (DSPs) and processors based on multicore processor architecture, as non-limiting examples. The apparatus 500 may have multiple processors, such as an application specific integrated circuit chip that is slaved in time to a clock which synchronizes the main processor.

[0070] Generally, various embodiments of the present disclosure may be implemented in hardware or special purpose circuits, software, logic or any combination thereof. Some aspects may be implemented in hardware, while other aspects may be implemented in firmware or software which may be executed by a controller, microprocessor or other computing device. While various aspects of embodiments of the present disclosure are illustrated and described as block diagrams, flowcharts, or using some other pictorial representation, it will be appreciated that the blocks, apparatus, systems, techniques or methods described herein may be implemented in, as non-limiting examples, hardware, software, firmware, special purpose circuits or logic, general purpose hardware or controller or other computing devices, or some combination thereof.

[0071] By way of example, embodiments of the present disclosure can be described in the general context of machine-executable instructions, such as those included in program modules, being executed in a device on a target real or virtual processor. Generally, program modules include routines, programs, libraries, objects, classes, components, data structures, or the like that perform particular tasks or implement particular abstract data types. The functionality of the program modules may be combined or split between program modules as desired in various embodiments. Machine-executable instructions for program modules may be executed within a local or distributed device. In a distributed device, program modules may be located in both local and remote storage media.

[0072] Program code for carrying out methods of the present disclosure may be written in any combination of one or more programming languages. These program codes may be provided to a processor or controller of a general purpose computer, special purpose computer, or other programmable data processing apparatus, such that the program codes, when executed by the processor or controller, cause the functions/operations specified in the flowcharts and/or block diagrams to be implemented. The program code may execute entirely on a machine, partly on the machine, as a stand-alone software package,

partly on the machine and partly on a remote machine or entirely on the remote machine or server.

[0073] In the context of this disclosure, a machine readable medium may be any tangible medium that may contain, or store a program for use by or in connection with an instruction execution system, apparatus, or device. The machine readable medium may be a machine readable signal medium or a machine readable storage medium. A machine readable medium may include but not limited to an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples of the machine readable storage medium would include an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing.

[0074] Further, while operations are depicted in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Likewise, while several specific implementation details are contained in the above discussions, these should not be construed as limitations on the scope of the present disclosure, but rather as descriptions of features that may be specific to particular embodiments. Certain features that are described in the context of separate embodiments may also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment may also be implemented in multiple embodiments separately or in any suitable sub-combination.

[0075] Although the present disclosure has been described in language specific to structural features and/or methodological acts, it is to be understood that the present disclosure defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

WHAT IS CLAIMED IS:

1. A method implemented at a pooling layer of a convolutional neural network, comprising:

receiving a channel including a two-dimensional array of elements, the elements being associated with a feature extracted from data being processed by the convolutional neural network;

dividing the channel into regions, each region including a sub-array of the array;

calculating magnitudes of Fourier transformed elements of each region; and

reshaping the calculated magnitudes of respective regions into respective vectors to complete pooling of the feature.

2. The method of claim 1, wherein the receiving comprises:

receiving the channel from a convolutional layer of the convolutional neural network.

3. The method of claim 1, wherein the dividing comprises:

dividing the channel into non-overlapping regions.

4. The method of claim 3, wherein each region has a same quantity of elements.

5. The method of claim 4, wherein each region has n×n elements, and n is an integer equal to or greater than two.

6. The method of claim 1, wherein the reshaping comprises:

arranging the calculated magnitudes along the channel dimension.

7. The method of claim 6, wherein the arranging comprises:

arranging the calculated magnitudes of the respective regions in a same order.

8. The method of claim 1, wherein the data is associated with at least one of image recognition, acoustic recognition, object detection, advanced driver assistance system,

self-driving cars, optical character recognition, face recognition, large-scale image classification, and human machine interaction.

9. An apparatus implemented at a pooling layer of a convolutional neural network, comprising:

at least one processor; and

at least one memory including computer program code; wherein

the at least one memory and the computer program code are configured to, with the at least one processor, cause the apparatus at least to:

receive a channel including a two-dimensional array of elements, the elements being associated with a feature extracted from data being processed by the convolutional neural network;

divide the channel into regions, each region including a sub-array of the array;

calculate magnitudes of Fourier transformed elements of each region; and

reshape the calculated magnitudes of respective regions into respective vectors to complete pooling of the feature.

10. The apparatus of claim 9, wherein the at least one memory and the computer program code are further configured to, with the at least one processor, cause the apparatus to:

receive the channel from a convolutional layer of the convolutional neural network.

11. The apparatus of claim 9, wherein the at least one memory and the computer program code are further configured to, with the at least one processor, cause the apparatus to:

divide the channel into non-overlapping regions.

12. The apparatus of claim 11, wherein each region has a same quantity of elements.

13. The apparatus of claim 12, wherein each region has n×n elements, and n is an

integer equal to or greater than two.

14. The apparatus of claim 9, wherein the at least one memory and the computer program code are further configured to, with the at least one processor, cause the apparatus to:

reshape the calculated magnitudes of respective regions by arranging the calculated magnitudes along the channel dimension.

15. The apparatus of claim 14, wherein the at least one memory and the computer program code are further configured to, with the at least one processor, cause the apparatus to:

arrange the calculated magnitudes of the respective regions in a same order.

16. The apparatus of claim 9, wherein the data is associated with at least one of image recognition, acoustic recognition, object detection, advanced driver assistance system, self-driving cars, optical character recognition, face recognition, large-scale image classification, and human machine interaction.

17. An apparatus comprising means for performing a method according to any of Claims 1 to 8.

18. A computer program product comprising at least one computer readable non-transitory memory medium having program code stored thereon, wherein the program code, when executed by an apparatus, causes the apparatus to perform a method according to any of Claims 1 to 8.

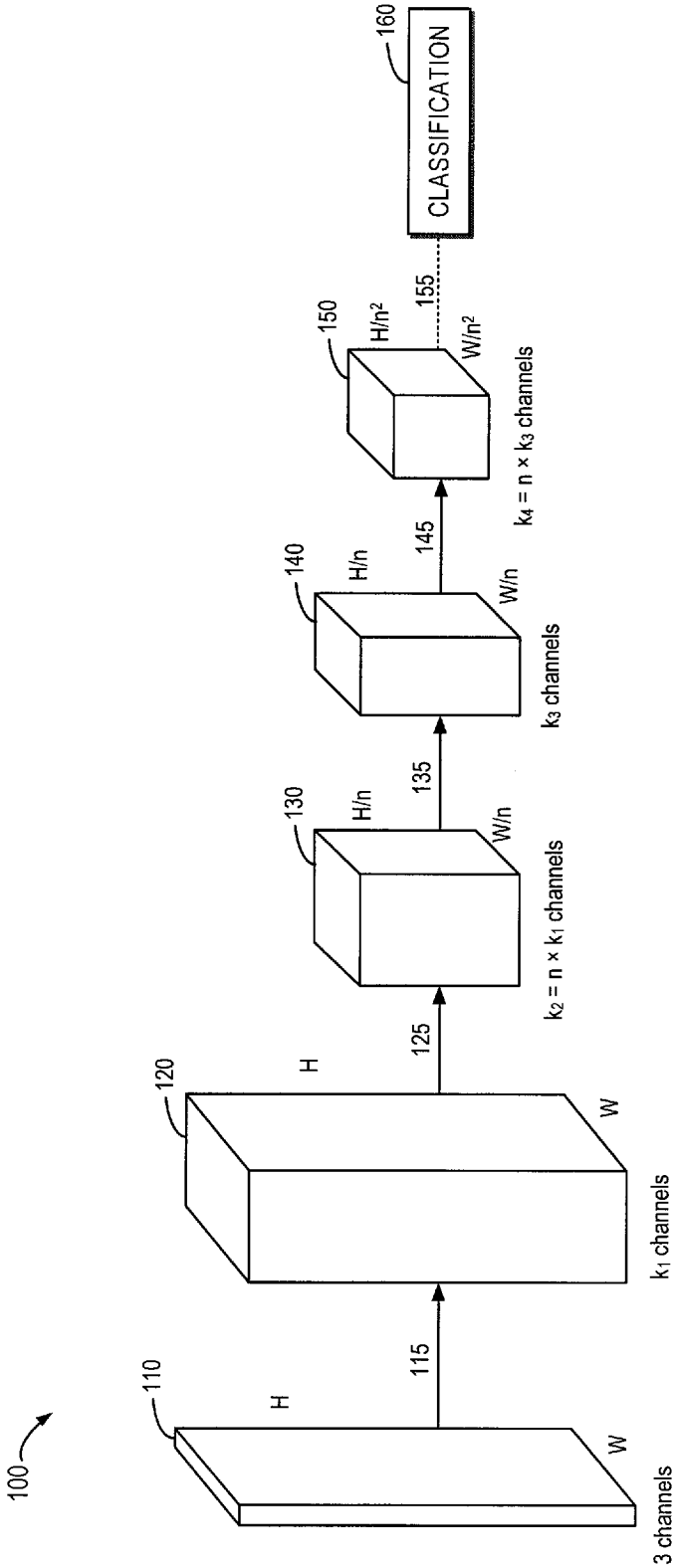
100
110
H
W
3 channels
115
120
H
W
k1 channels
125
130
H/n
W/n
k2 = n × k1 channels
135
140
H/n
W/n
k3 channels
145
150
H/n²
W/n²
k4 = n × k3 channels
155
160
CLASSIFICATION

Fig. 1

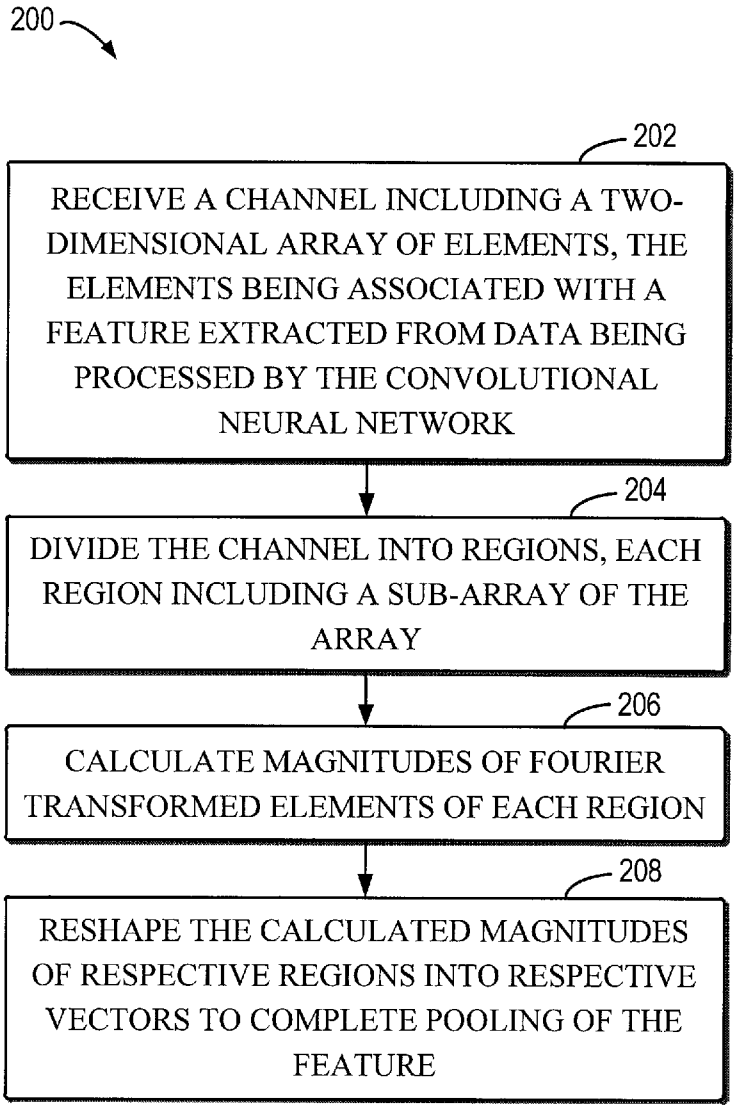
200
202
RECEIVE A CHANNEL INCLUDING A TWO-DIMENSIONAL ARRAY OF ELEMENTS, THE ELEMENTS BEING ASSOCIATED WITH A FEATURE EXTRACTED FROM DATA BEING PROCESSED BY THE CONVOLUTIONAL NEURAL NETWORK
204
DIVIDE THE CHANNEL INTO REGIONS, EACH REGION INCLUDING A SUB-ARRAY OF THE ARRAY
206
CALCULATE MAGNITUDES OF FOURIER TRANSFORMED ELEMENTS OF EACH REGION
208
RESHAPE THE CALCULATED MAGNITUDES OF RESPECTIVE REGIONS INTO RESPECTIVE VECTORS TO COMPLETE POOLING OF THE FEATURE

Fig. 2

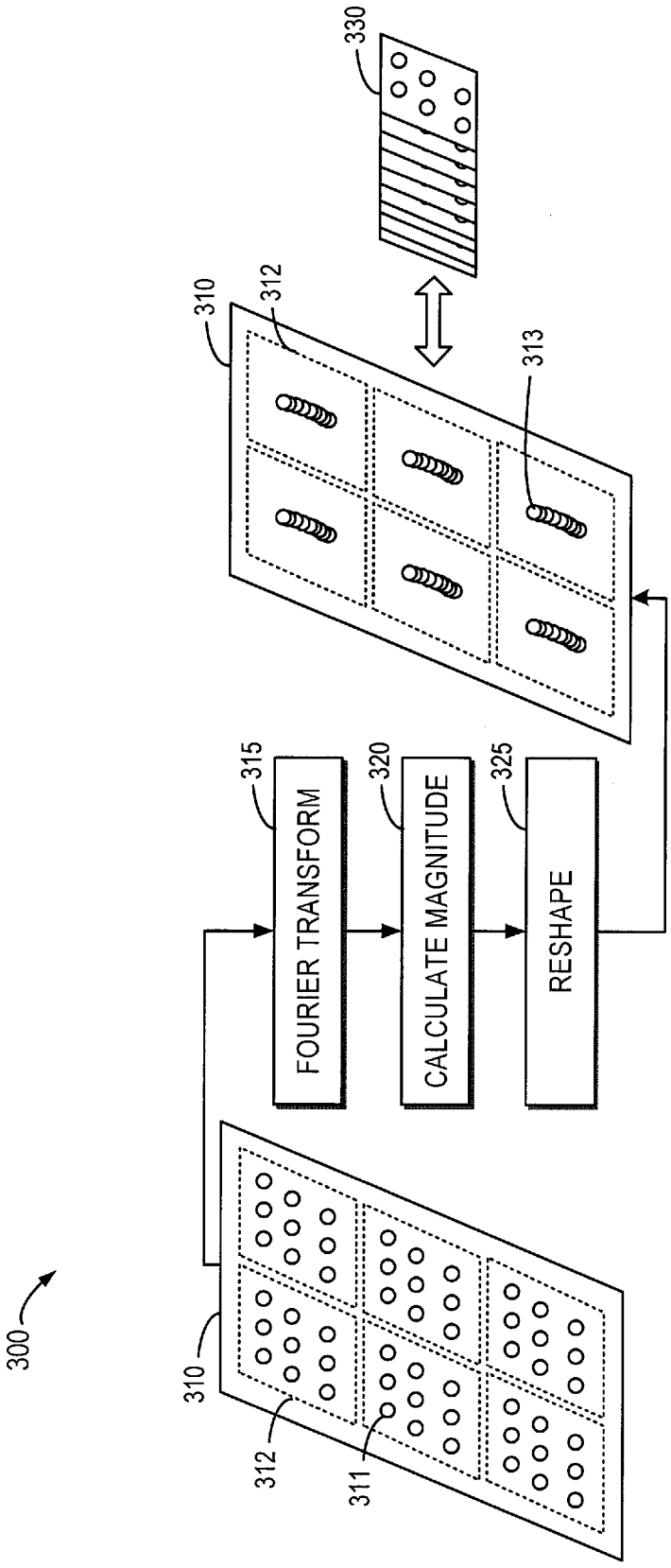
300
310
312
311
315
FOURIER TRANSFORM
320
CALCULATE MAGNITUDE
325
RESHAPE
310
312
313
330

Fig. 3

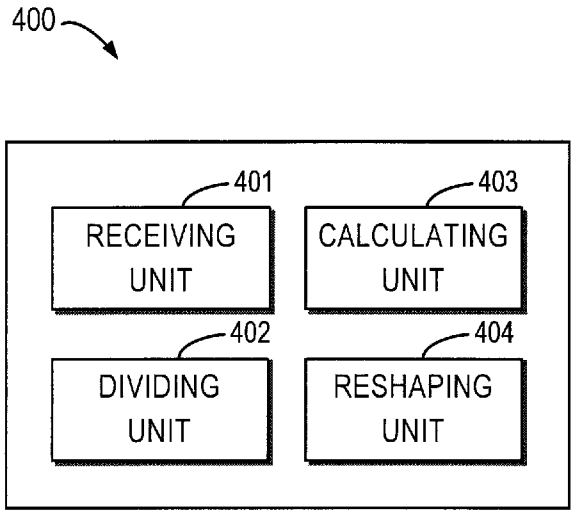
400
401
RECEIVING UNIT
403
CALCULATING UNIT
402
DIVIDING UNIT
404
RESHAPING UNIT

Fig. 4

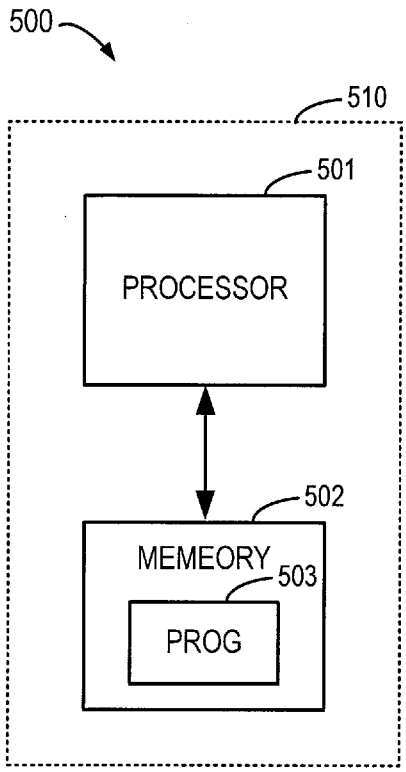

Fig. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/086152

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i

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)

G06F, G06K

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)

CNABS,CNKI,DWPI,SIPOABS: neural network, convolutional, convolution, pooling layer, divide, channel, magnitude, reshape, vector, array

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2014288928 A1 (PENN, G.B. ET AL.) 25 September 2014 (2014-09-25) the whole document	1-18
A	US 2016104056 A1 (MICROSOFT TECHNOLOGY LICENSING LLC) 14 April 2016 (2016-04-14) the whole document	1-18
A	CN 104463172 A (CHINESE ACAD SCI CHONGQING GREEN & INTEL) 25 March 2015 (2015-03-25) the whole document	1-18

☐ 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

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 March 2017	**24 March 2017**

Name and mailing address of the ISA/CN	Authorized officer
STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088** **China**	**DENG,Jun**
Facsimile No. **(86-10)62019451**	Telephone No. **(86-10)62411644**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/086152

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 2014288928 A1	25 September 2014	US 9190053 B2	17 November 2015
US 2016104056 A1	14 April 2016	WO 2016054779 A1	14 April 2016
		CN 105917354 A	31 August 2016
		US 9542621 B2	10 January 2017
CN 104463172 A	25 March 2015	None	