



(51) International Patent Classification:

G06N 3/08 (2006.01)

G06N 3/04 (2006.01)

(21) International Application Number:

PCT/US2019/048722

(22) International Filing Date:

29 August 2019 (29.08.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/727,730 06 September 2018 (06.09.2018) US

62/878,777 26 July 2019 (26.07.2019) US

16/535,681 08 August 2019 (08.08.2019) US

(71) Applicant: NEC LABORATORIES AMERICA, INC.

[US/US]; 4 Independence Way, Suite 200, Princeton, NJ 08540 (US).

(72) Inventors: TRAN, Quoc-Huy; 2425 Karen Drive, Santa

Clara, CA 95050 (US). TSAI, Yi-Hsuan; 1070 Universi-

ty Avenue, San Jose, CA 95126 (US). SOHN, Kihyuk; 37168 Meadowbrook Common, Fremont, CA 94536 (US). LIU, Buyu; 11527 Bianchini Lane, Cupertino, CA 95014 (US). CHANDRAKER, Manmohan; 444 Saratoga Avenue, Santa Clara, CA 95050 (US). SU, Jong-Chyi; 36 Hunters Hill Circle, Amherst, MA 01002 (US).

(74) Agent: BITETTO, James J.; Tutunjian & Bitetto, P.C., 401 Broadhollow Road, Suite 402, Melville, NY 11747 (US).

(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, 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,

(54) Title: DOMAIN ADAPTATION FOR INSTANCE DETECTION AND SEGMENTATION

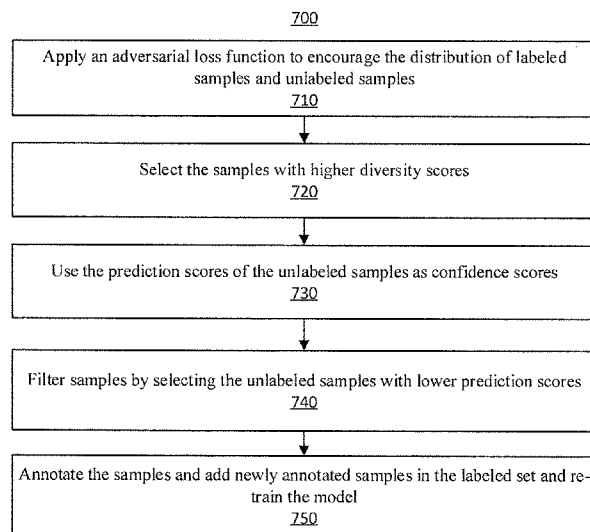


FIG. 5

(57) Abstract: Systems and methods for domain adaptation are provided. The system aligns image level features between a source domain and a target domain based on an adversarial learning process while training a domain discriminator (710). The system selects, using the domain discriminator, unlabeled samples from the target domain that are far away from existing annotated samples from the target domain (720). The system selects, based on a prediction score of each of the unlabeled samples (730), samples with lower prediction scores (740). The system annotates the samples with the lower prediction scores (750).



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).

Published:

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

DOMAIN ADAPTATION FOR INSTANCE DETECTION AND SEGMENTATION

RELATED APPLICATION INFORMATION

[0001] This application claims priority to U.S. Provisional Patent Application No. 62/727,730, filed on September 6, 2018, U.S. Provisional Patent Application No. 62/878,777, filed on July 26, 2019, and U.S. Patent Application No. 16/535,681, filed on August 8, 2019, incorporated herein by reference in their entirety.

BACKGROUND

Technical Field

[0002] The present invention relates to domain management and more particularly to transferring information between domains.

Description of the Related Art

[0003] Road-scene images from two cities may vary significantly due to the different style of architectures and roads. In addition, images may be captured under various conditions such as weather and lighting environment even within the same city. Therefore, instance object detection and segmentation of images is known to require labor-intensive annotation process to train the fully supervised model in every domain.

SUMMARY

[0004] According to an aspect of the present invention, a method is provided for domain adaptation. The method includes applying an adversarial loss function to encourage a distribution of labeled samples and unlabeled samples from a label set. The method also includes selecting higher diversity score unlabeled samples from the unlabeled samples. The method includes selecting lower prediction score higher

diversity score unlabeled samples from the higher diversity score unlabeled samples, and annotating the lower prediction score higher diversity score unlabeled samples. The method further includes updating the label set with annotated lower prediction score higher diversity score unlabeled samples to form an updated labeled set, and retraining a model using the updated label set.

[0005] According to another aspect of the present invention, a system is provided for domain adaptation. The system aligns image level features between a source domain and a target domain based on an adversarial learning process while training a domain discriminator. The system selects, using the domain discriminator, unlabeled samples from the target domain that are far away from existing annotated samples from the target domain. The system selects, based on a prediction score of each of the unlabeled samples, samples with lower prediction scores. The system annotates the samples with the lower prediction scores.

[0006] These and other features and advantages will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

[0007] The disclosure will provide details in the following description of preferred embodiments with reference to the following figures wherein:

[0008] FIG. 1 is a generalized diagram of a neural network, in accordance with an embodiment of the present invention;

[0009] FIG. 2 is a diagram of an artificial neural network (ANN) architecture, in accordance with an embodiment of the present invention;

[0010] FIG. 3 is a block diagram illustrating a high-level system for semi-supervised domain adaptation for instance detection and segmentation, in accordance with an embodiment of the present invention;

[0011] FIG. 4 is a block diagram illustrating an implementation of semi-supervised domain adaptation for instance detection and segmentation, in accordance with an embodiment of the present invention;

[0012] FIG. 5 is a flow diagram illustrating a method for semi-supervised domain adaptation for instance detection and segmentation, in accordance with the present invention; and

[0013] FIG. 6 is a block diagram illustrating an implementation of a system for semi-supervised domain adaptation for instance detection and segmentation, in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0014] In accordance with embodiments of the present invention, systems and methods are provided for domain adaptation for (object) instance detection and segmentation. The system applies an adversarial loss function to encourage a distribution of labeled samples and unlabeled samples from a label set. The system selects higher diversity score unlabeled sample from the unlabeled samples. The higher diversity score unlabeled samples are unlabeled samples with higher diversity scores than other unlabeled samples. The system selects lower prediction score higher diversity score unlabeled samples from the higher diversity score unlabeled samples. The lower prediction score higher diversity score unlabeled samples are higher diversity score unlabeled samples with lower prediction scores than other higher diversity score unlabeled samples. The system annotates the lower prediction score higher diversity

score unlabeled samples. The system updates the label set with annotated lower prediction score higher diversity score unlabeled samples to form an updated labeled set, and retrain a model using the updated label set. The model implements at least one predetermined task, for example, object detection and semantic segmentation.

[0015] In one embodiment, the system provides an active learning framework by using a domain discriminator to select unlabeled images that are far away from existing annotated images. As the domain adaptation and active learning processes continue, the domain discriminator becomes stronger to ensure that the selected images are diverse for annotations. In addition, the system combines a classifier score to further filter these images that have less confidence for prediction. After selecting the target images, the system annotates the selected target images and retrain the model in an iterative scheme.

[0016] Embodiments described herein may be entirely hardware, entirely software or including both hardware and software elements. In a preferred embodiment, the present invention is implemented in software, which includes but is not limited to firmware, resident software, microcode, etc.

[0017] Embodiments may include a computer program product accessible from a computer-usable or computer-readable medium providing program code for use by or in connection with a computer or any instruction execution system. A computer-usable or computer readable medium may include any apparatus that stores, communicates, propagates, or transports the program for use by or in connection with the instruction execution system, apparatus, or device. The medium can be magnetic, optical, electronic, electromagnetic, infrared, or semiconductor system (or apparatus or device) or a propagation medium. The medium may include a computer-readable storage medium such as a semiconductor or solid state memory, magnetic tape, a removable

computer diskette, a random access memory (RAM), a read-only memory (ROM), a rigid magnetic disk and an optical disk, etc.

[0018] Each computer program may be tangibly stored in a machine-readable storage media or device (e.g., program memory or magnetic disk) readable by a general or special purpose programmable computer, for configuring and controlling operation of a computer when the storage media or device is read by the computer to perform the procedures described herein. The inventive system may also be considered to be embodied in a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

[0019] A data processing system suitable for storing and/or executing program code may include at least one processor coupled directly or indirectly to memory elements through a system bus. The memory elements can include local memory employed during actual execution of the program code, bulk storage, and cache memories which provide temporary storage of at least some program code to reduce the number of times code is retrieved from bulk storage during execution. Input/output or I/O devices (including but not limited to keyboards, displays, pointing devices, etc.) may be coupled to the system either directly or through intervening I/O controllers.

[0020] Network adapters may also be coupled to the system to enable the data processing system to become coupled to other data processing systems or remote printers or storage devices through intervening private or public networks. Modems, cable modem and Ethernet cards are just a few of the currently available types of network adapters.

[0021] Referring now to FIG. 1, a generalized diagram of a neural network is shown.

[0022] An artificial neural network (ANN) is an information processing system that is inspired by biological nervous systems, such as the brain. The key element of ANNs is the structure of the information processing system, which includes many highly interconnected processing elements (called “neurons”) working in parallel to solve specific problems. ANNs are furthermore trained in-use, with learning that involves adjustments to weights that exist between the neurons. An ANN is configured for a specific application, such as pattern recognition or data classification, through such a learning process.

[0023] ANNs demonstrate an ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be detected by humans or other computer-based systems. The structure of a neural network generally has input neurons 102 that provide information to one or more “hidden” neurons 104. Connections 108 between the input neurons 102 and hidden neurons 104 are weighted and these weighted inputs are then processed by the hidden neurons 104 according to some function in the hidden neurons 104, with weighted connections 108 between the layers. There can be any number of layers of hidden neurons 104, and as well as neurons that perform different functions. There exist different neural network structures as well, such as convolutional neural network, maxout network, etc. Finally, a set of output neurons 106 accepts and processes weighted input from the last set of hidden neurons 104.

[0024] This represents a “feed-forward” computation, where information propagates from input neurons 102 to the output neurons 106. Upon completion of a feed-forward computation, the output is compared to a desired output available from training data. The error relative to the training data is then processed in “feed-back” computation, where the hidden neurons 104 and input neurons 102 receive information regarding the

error propagating backward from the output neurons 106. Once the backward error propagation has been completed, weight updates are performed, with the weighted connections 108 being updated to account for the received error. This represents just one variety of ANN.

[0025] As described herein, the ANN 100 can implement instance object detection and segmentation based on input images, such as described below with respect to FIG. 6.

[0026] Referring now to the drawings in which like numerals represent the same or similar elements and initially to FIG. 2, an artificial neural network (ANN) architecture 200 is shown. It should be understood that the present architecture is purely exemplary and that other architectures or types of neural network may be used instead. The ANN embodiment described herein is included with the intent of illustrating general principles of neural network computation at a high level of generality and should not be construed as limiting in any way.

[0027] Furthermore, the layers of neurons described below and the weights connecting them are described in a general manner and can be replaced by any type of neural network layers with any appropriate degree or type of interconnectivity. For example, layers can include convolutional layers, pooling layers, fully connected layers, stopmax layers, or any other appropriate type of neural network layer. Furthermore, layers can be added or removed as needed and the weights can be omitted for more complicated forms of interconnection.

[0028] During feed-forward operation, a set of input neurons 202 each provide an input signal in parallel to a respective row of weights 204. In the hardware embodiment described herein, the weights 204 each have a respective settable value, such that a weight output passes from the weight 204 to a respective hidden neuron 206 to represent

the weighted input to the hidden neuron 206. In software embodiments, the weights 204 may simply be represented as coefficient values that are multiplied against the relevant signals. The signals from each weight adds column-wise and flows to a hidden neuron 206.

[0029] The hidden neurons 206 use the signals from the array of weights 204 to perform some calculation. The hidden neurons 206 then output a signal of their own to another array of weights 204. This array performs in the same way, with a column of weights 204 receiving a signal from their respective hidden neuron 206 to produce a weighted signal output that adds row-wise and is provided to the output neuron 208.

[0030] It should be understood that any number of these stages may be implemented, by interposing additional layers of arrays and hidden neurons 206. It should also be noted that some neurons may be constant neurons 209, which provide a constant output to the array. The constant neurons 209 can be present among the input neurons 202 and/or hidden neurons 206 and are only used during feed-forward operation.

[0031] During back propagation, the output neurons 208 provide a signal back across the array of weights 204. The output layer compares the generated network response to training data and computes an error. The error signal can be made proportional to the error value. In this example, a row of weights 204 receives a signal from a respective output neuron 208 in parallel and produces an output which adds column-wise to provide an input to hidden neurons 206. The hidden neurons 206 combine the weighted feedback signal with a derivative of its feed-forward calculation and stores an error value before outputting a feedback signal to its respective column of weights 204. This back-propagation travels through the entire network 200 until all hidden neurons 206 and the input neurons 202 have stored an error value.

[0032] During weight updates, the stored error values are used to update the settable values of the weights 204. In this manner the weights 204 can be trained to adapt the neural network 200 to errors in its processing. It should be noted that the three modes of operation, feed forward, back propagation, and weight update, do not overlap with one another.

[0033] A convolutional neural networks (CNN) is a subclass of ANNs which has at least one convolution layer. A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN consist of convolutional layers, rectified linear unit (RELU) layer (e.g., activation function), pooling layers, fully connected layers and normalization layers. Convolutional layers apply a convolution operation to the input and pass the result to the next layer. The convolution emulates the response of an individual neuron to visual stimuli.

[0034] CNNs can be applied to analyzing visual imagery. CNNs can capture local information (e.g., neighbor pixels in an image or surrounding words in a text) as well as reduce the complexity of a model (to allow, for example, faster training, requirement of fewer samples, and reduction of the chance of overfitting).

[0035] CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. CNNs are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. CNNs can be used for applications in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing. CNNs can be applied to implement instance object detection and segmentation based on input images, such as described herein below with respect to FIG. 3. The CNNs can also apply adversarial loss functions, as further described herein below with respect to FIGS. 3 and 4.

[0036] Referring now to FIG. 3, a high-level system 300 for semi-supervised domain adaptation for instance detection and segmentation is illustratively depicted in accordance with an embodiment of the present invention.

[0037] In a system with semi-supervised domain adaptation, the system is able to acquire annotations for some (for example, a few) images from the target domain (and thereby provide a (for example, limited) measure of supervision). As shown in FIG. 3, system 300 includes main task system 305, a domain adaptation system 400 and an active learning system 500. Note that boxes with square edges denote data (for example, source (I_{src}), labeled target (I_{tar_l}), etc.) or computational units (or devices), such as convolutional neural networks (CNNs) or clustering. Rounded edge boxes denote loss functions (for example, task loss 375 and adversarial loss 405). Arrows indicate data flow.

[0038] The combination of the domain adaptation system 400 and the active learning system 500 provide a framework for jointly considering the diversity and uncertainty that are estimated by main task convolutional neural networks (CNNs) 310. The domain discriminator 410 implements domain transfer (for example, shift of knowledge with annotations) and provides a signal to efficiently select informative unlabeled target samples. By jointly considering the diversity and uncertainty that are estimated from the main task CNNs 310, the framework (provided by the domain adaptation system 400 and the active learning system 500) iteratively updates both the model prediction (via main task CNNs 310) and active learning (via the active learning system 500). Consequently, the example embodiments provide a framework can achieve higher accuracy while annotating fewer samples. Note that the framework is general and can be also applied for various tasks (for example, semantic segmentation and fine-grained

object recognition). The example embodiments achieve higher accuracy for object detection based on a limited budget for annotations in the target domain.

[0039] Main task CNNs 310, as shown in FIG. 3, receive input images. Main task CNNs 310 implement instance object detection and segmentation based on input images. The input images can come from two domains, denoted herein as source (I_{src} 315), labeled target (I_{tar_l} 320), and unlabeled target (I_{tar_u} 325). These inputs are then fed into the image-level CNN (IL CNN 330) to generate features (F_{src} 335, F_{tar_l} 340, and F_{tar_u} 345). A feature can include an edge, corners, interest points, blobs, regions of interest points, ridges, etc. Based on these features, the main task CNNs 310 propagate the features to the region of interest level (ROI-level) CNN (denoted in FIG. 3 as ROI_L CNN 350) and generate output bounding boxes (O_{src} 355, O_{tar_l} 360, and O_{tar_u} 365) as the (for example, final) object detection results. A bounding box is the coordinates of the rectangular border that fully encloses a digital image when it is placed over a page, a canvas, a screen or other similar (for example, bi-dimensional) background. Note that main task CNNs 310 can also predict an instance segmentation map within each bounding box. Instance segmentation determines a pixel-wise mask for each object in an image even if the objects are of the same class label.

[0040] Main task loss function 370 trains the CNNs from main task CNNs 310. For images from the source domain and the labeled target domain, main task loss function 370 accesses (for example, is given, receives, etc.) ground truth labels (GT_{src} 380 and GT_{tar_l} 385), which are used in a standard supervised loss function (task loss 375) to train the main task CNNs 310. Ground truth refers to the accuracy of the training set's classification for supervised learning techniques and is used in statistical models by the main task CNNs 310 to prove or disprove hypotheses regarding images from the source

domain. A loss function measures the inconsistency between a predicted value and an actual label.

[0041] Domain adaptation system 400 adapts domains (via a domain discriminator). In order to train the main task CNNs 310 and also handle images from the unlabeled target domain (I_{tar_u} 325), domain adaptation system 400 applies an adversarial loss 405 function (or regularization) to encourage the distribution of labeled samples (F_{src} 335 and F_{tar_l} 340) and unlabeled samples (F_{tar_u}) to be similar. This adversarial loss 405 function is computed based on a domain discriminator 410 that aims to distinguish between labeled and unlabeled samples for the domain alignment purpose. Domain adaptation system 400 solves (or resolves, etc.) the domain gap (for example, problem, issue, etc.) between two domains via aligning the image-level features (F_{src} 335 and F_{tar_l} 340) using an adversarial learning scheme. During this process, a domain discriminator 410 is adopted to distinguish whether the feature is from the source (labeled, for example, as 0) or the target (labeled, for example, as 1) domain.

[0042] When introducing a limited budget for annotations in the target domain, system 300 aims to select informative images that can improve the target performance efficiently, for example, annotating fewer images while achieving higher performance. Toward this end, system 300 implements an active learning framework by using the domain discriminator 410 to select unlabeled images that are far away from existing annotated images. As the domain adaptation and active learning processes continue, the domain discriminator 410 becomes stronger to ensure that the selected images are diverse for annotations.

[0043] Active learning system 500 implements active learning via sample selection. Given a limited budget to annotate unlabeled target images, active learning system 500 iteratively selects informative samples from the target domain for annotations. First,

active learning system 500 uses the domain discriminator 410 from domain adaptation system 400 as an indicator to find a set of unlabeled target images that are diverse from the existing labeled samples. To achieve this, active learning system 500 applies the domain discriminator 410 on all the unlabeled samples (I_tar_u 325) and ranks the scores of the unlabeled samples. By selecting the unlabeled samples (I_tar_u 325) with higher scores, for example, far away from the labeled samples in the feature space, active learning system 500 ensures that these selected samples are different from existing labeled samples and thereby increases diversity among the samples.

[0044] After selecting unlabeled samples that are different from existing labeled samples as described above, active learning system 500 further uses their prediction scores (O_tar_u 365) from the main task CNNs 310 as confidence scores. Active learning system 500 then filter samples by selecting the samples with lower scores to account for the uncertainty.

[0045] By using these two criteria to select unlabeled samples (I_tar_u 325), active learning system 500 annotates the unlabeled samples and adds these newly annotated samples in the labeled set (I_tar_l 320) and re-trains the model. Active learning system 500 iteratively repeats this process by training CNNs (for example, main task CNNs 310) and the domain discriminator 410, in a manner that the diversity (for example, based on I_tar_u 325) and uncertainty (for example, confidence based on O_tar_u 365) are considered in each iteration. In other words, active learning system 500 combines the classifier score to further filter images that have less confidence for prediction. After selecting the target images, active learning system 500 annotates the selected target images and re-trains the model in an iterative scheme.

[0046] The domain discriminator 410 provides a cue for the sample selector 510 to better understand the distribution difference between labeled and unlabeled samples.

Consequently, system 300 uses the cue in a key manner to obtain a small and informative set of samples while achieving better task performance (for example, than incumbent processes). Sample selector 510 provides (for example, sends 515) the sample annotation 515 to I_tar_l 320 and I_tar_u 325 (for example, as sample updates).

[0047] Referring now to FIG. 4, an implementation of semi-supervised domain adaptation for instance detection and segmentation is illustratively depicted in accordance with an embodiment of the present invention.

[0048] As shown in FIG. 4, the example embodiments are directed towards (for example, target) instance object detection and segmentation, which in incumbent implementations is known to require labor-intensive annotation process to train the fully supervised model in every domain. In contrast, the example embodiments address annotation bottlenecks by transferring the knowledge from one domain with full annotations to another domain with only a limited budget to annotate the ground truth for training the models.

[0049] For example, with respect to implementation 600, using similar methods as described with respect to system 300, learning module 615 can learn an object detection (for example, car, bike, and pedestrian) model using a large amount of existing data 610 (for example, training data) with annotations in the source domain 605 (for example, daytime images for daytime object detection 620). For the nighttime case (nighttime object detection 665), which can be more challenging than the daytime case, the sample selector 510 can use the daytime model to select informative data (new data 660) from the nighttime domain 655 for annotations (with a limited budget) and re-train the object detector (for example, nighttime object detection 665) to improve the performance during nighttime.

[0050] The example embodiments exploit a few images with annotations in the target domain, which is a practical setting when there is a limited budget available for annotations. The example embodiments provide an active learning technique for the semi-supervised domain adaptation for example, in instances in which only a limited budget available to annotate the target images, e.g., limited time and money.

[0051] The example embodiments provide a general framework that can be applied for various types of domains and tasks, for example for labor-intensive tasks such as object detection, semantic segmentation, and fine-grained object recognition.

[0052] Referring now to FIG. 5, a method 700 for semi-supervised domain adaptation for instance detection and segmentation is illustratively depicted in accordance with an embodiment of the present invention.

[0053] At block 710, system 300 applies an adversarial loss function to encourage the distribution of labeled samples and unlabeled samples. The samples are from a target domain. The system 300 also receives input samples (for example, images, videos, or other data forms such as texts) from a source domain. The source domain and the target domain can include images taken under different conditions, such as scene images from two cities, which vary significantly due to the different style of architectures and roads, two different geographical areas, two different lighting conditions, etc.

[0054] At block 720, system 300 selects the samples with higher diversity scores, (for example, far away from the labeled samples in the feature space) to ensure that these selected unlabeled samples are different from existing labeled samples.

[0055] At block 730, system 300 uses the prediction scores of the selected unlabeled samples from the main task CNNs 310 as confidence scores. System 300 then filters

samples by selecting the unlabeled samples with lower prediction scores to account for the uncertainty (at block 740).

[0056] System 300 uses these two criteria (higher diversity scores and lower prediction scores) to select unlabeled samples. At block 750, system 300 annotates the samples and adds these newly annotated samples in the labeled set and re-trains the model. System 300 iteratively repeats this process by training CNNs and the domain discriminator, and thereby ensures that the diversity and uncertainty are considered in each iteration.

[0057] Referring now to FIG. 6, an implementation of a device that includes a system for semi-supervised domain adaptation for instance detection and segmentation is illustratively depicted in accordance with an embodiment of the present invention.

[0058] As shown in FIG. 6, device 800 can include a camera or other image processing device that uses semi-supervised domain adaptation for instance detection and segmentation. According to an example embodiment, the semi-supervised domain adaptation for instance detection and segmentation can include an automatic process in which a camera is continuously collecting target images and selecting which data to annotate. The selected images can be sent to an associated device (or interface, etc.) for manual or semi-automatic annotation. Finally, annotated images are sent back to the semi-supervised domain adaptation device to automatically finetune the model. These processes can be parallelly and iteratively performed. For example, using similar methods as described with respect to system 300, learning module 815 can learn an object detection model using a large amount of source domain data 810 with annotations in the source domain 805 (for example, images detected with source conditions (for example, particular weather, geographic, etc.) for source condition object detection 820). The source condition and the target condition can be related (for

example, different weather condition) in a way that object detection and annotation is more easily accomplished in the source domain.

[0059] For target condition object detection 865, which can be more challenging than the daytime case, the sample selector 510 can use the source domain model to select informative data (target domain data 860) from the target domain 855 for annotations (with a limited budget) and re-train the object detector (for example, target domain object detection 865) to improve the performance when the target conditions are present. Later, the target condition object detection 865 can receive images captured under the target conditions. The object detector (target domain object detection 865) can detect and annotate images 870 (for example, from the target domain data 860) based on the training from learning module 815.

[0060] Thus, the example embodiments can assist systems that use imaging data from different domains, such as advanced driver-assistance systems (ADAS) and fine-grained object recognition systems, in an active learning fashion that continuously learns from the new data. Similarly, other scenarios such as weather conditions, lighting conditions and cross-city cases can be analyzed and transfer of knowledge from one domain with full annotations to another domain with only a limited budget can be inexpensively (and/or cost effectively) implemented by applying the same methodology.

[0061] The foregoing is to be understood as being in every respect illustrative and exemplary, but not restrictive, and the scope of the invention disclosed herein is not to be determined from the Detailed Description, but rather from the claims as interpreted according to the full breadth permitted by the patent laws. It is to be understood that the embodiments shown and described herein are only illustrative of the present invention and that those skilled in the art may implement various modifications without

departing from the scope and spirit of the invention. Those skilled in the art could implement various other feature combinations without departing from the scope and spirit of the invention. Having thus described aspects of the invention, with the details and particularity required by the patent laws, what is claimed and desired protected by Letters Patent is set forth in the appended claims.

WHAT IS CLAIMED IS:

1. A method for domain adaptation, comprising:
 - aligning image level features between a source domain and a target domain based on an adversarial learning process while training a domain discriminator (710);
 - selecting, using the domain discriminator, unlabeled samples from the target domain that are furthest away from existing annotated samples from the target domain (720);
 - selecting, by a processor device, based on a prediction score of each of the unlabeled samples (730), samples with lower prediction scores (740); and
 - annotating the samples with the lower prediction scores (750).
2. The method as recited in claim 1, further comprising:
 - iteratively retraining a model that annotates the unlabeled samples based on the annotated samples with the lower prediction scores, wherein the model implements at least one predetermined task.
3. The method as recited in claim 2, wherein the at least one predetermined task includes at least one of instance object detection and segmentation.
4. The method as recited in claim 2, wherein retraining the model further comprises:
 - inputting an updated label set including the annotated samples with the lower prediction scores into an image-level convolutional neural network (CNN) to generate at least one feature;

based on the at least one feature, propagating the updated label set to a region of interest level (ROI-level) CNN; and

generating output bounding boxes as at least one object detection.

5. The method as recited in claim 6, further comprising:

predicting an instance segmentation map within each bounding box.

6. The method as recited in claim 1, wherein aligning the image level features between the source domain and the target domain based on the adversarial learning process further comprises:

applying an adversarial loss function to encourage a distribution of labeled samples and the unlabeled samples from a label set;

selecting, by the processor device, at least one higher diversity score unlabeled sample from the unlabeled samples; and

selecting at least one lower prediction score higher diversity score unlabeled sample from the at least one higher diversity score unlabeled sample.

7. The method as recited in claim 6, further comprising:

annotating the at least one lower prediction score higher diversity score unlabeled sample; and

updating the label set with at least one annotated lower prediction score higher diversity score unlabeled sample to form an updated labeled set.

8. The method as recited in claim 6, wherein selecting the at least one lower prediction score higher diversity score unlabeled sample from the unlabeled samples further comprises:

using prediction scores of the unlabeled samples as confidence scores.

9. The method as recited in claim 1, wherein the source domain and the target domain are selected from at least one of different geographical areas, different weather conditions and different lighting conditions.

10. The method as recited in claim 1, wherein selecting the at least one higher diversity score unlabeled sample from the unlabeled samples further comprises:

selecting unlabeled images that are furthest away from existing annotated images in the label set.

11. The method as recited in claim 1, further comprising:

using a supervised loss function and ground truth labels from the source domain and the target domain to train at least one image-level convolutional neural network (CNN).

12. A computer system for domain adaptation, comprising:

a processor device operatively coupled to a memory device, the processor device being configured to:

align image level features between a source domain and a target domain based on an adversarial learning process while training a domain discriminator (710);

select, using the domain discriminator, unlabeled samples from the target domain that are far away from existing annotated samples from the target domain (720);

select based on a prediction score of each of the unlabeled samples (730), samples with lower prediction scores (740); and

annotate the samples with the lower prediction scores (750).

13. The system as recited in claim 12, wherein the processor device is further configured to:

iteratively retrain a model that annotates the unlabeled samples based on the annotated samples with the lower prediction scores, wherein the model implements at least one predetermined task.

14. The system as recited in claim 13, wherein the at least one predetermined task includes at least one of instance object detection and segmentation.

15. The system as recited in claim 13, wherein, when retraining the model, the processor device is further configured to:

input an updated label set including the annotated samples with the lower prediction scores into an image-level convolutional neural network (CNN) to generate at least one feature;

based on the at least one feature, propagate the updated label set to a region of interest level (ROI-level) CNN; and

generate output bounding boxes as at least one object detection.

16. The system as recited in claim 15, wherein the processor device is further configured to:

predict an instance segmentation map within each bounding box.

17. The system as recited in claim 13, wherein, when aligning the image level features between the source domain and the target domain based on the adversarial learning process, the processor device is further configured to:

apply an adversarial loss function to encourage a distribution of labeled samples and the unlabeled samples from a label set;

select at least one higher diversity score unlabeled sample from the unlabeled samples; and

selecting at least one lower prediction score higher diversity score unlabeled sample from the at least one higher diversity score unlabeled sample.

18. The system as recited in claim 12, wherein the source domain and the target domain are selected from at least one of different geographical areas, different weather conditions and different lighting conditions.

19. The system as recited in claim 12, wherein the processor device is further configured to:

use a supervised loss function and ground truth labels from the source domain and the target domain to train at least one image-level convolutional neural network (CNN).

20. A computer program product for domain adaptation, the computer program product comprising a non-transitory computer readable storage medium having program instructions embodied therewith, the program instructions executable by a computing device to cause the computing device to perform the method comprising:

aligning image level features between a source domain and a target domain based on an adversarial learning process while training a domain discriminator (710);

selecting, using the domain discriminator, unlabeled samples from the target domain that are far away from existing annotated samples from the target domain (720);

selecting, by a processor device, based on a prediction score of each of the unlabeled samples (730), samples with lower prediction scores (740); and

annotating the samples with the lower prediction scores (750).

100

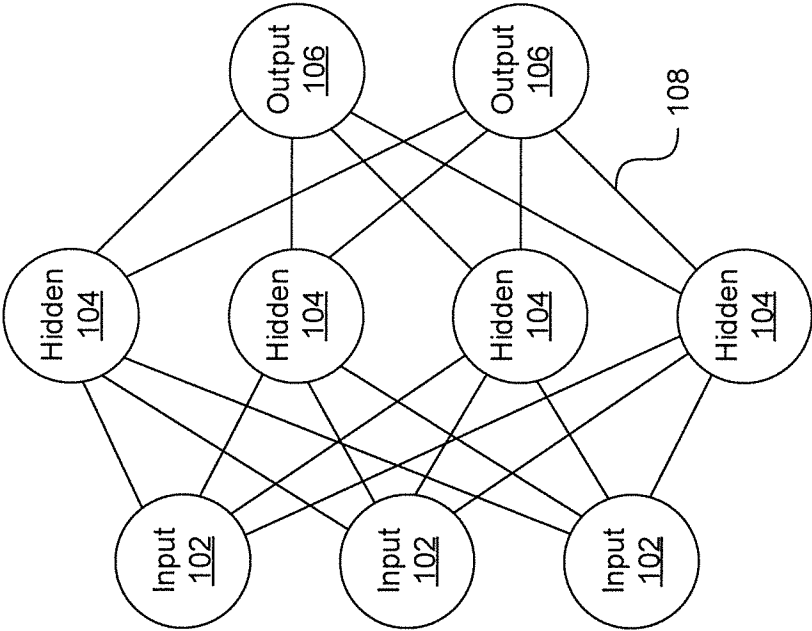
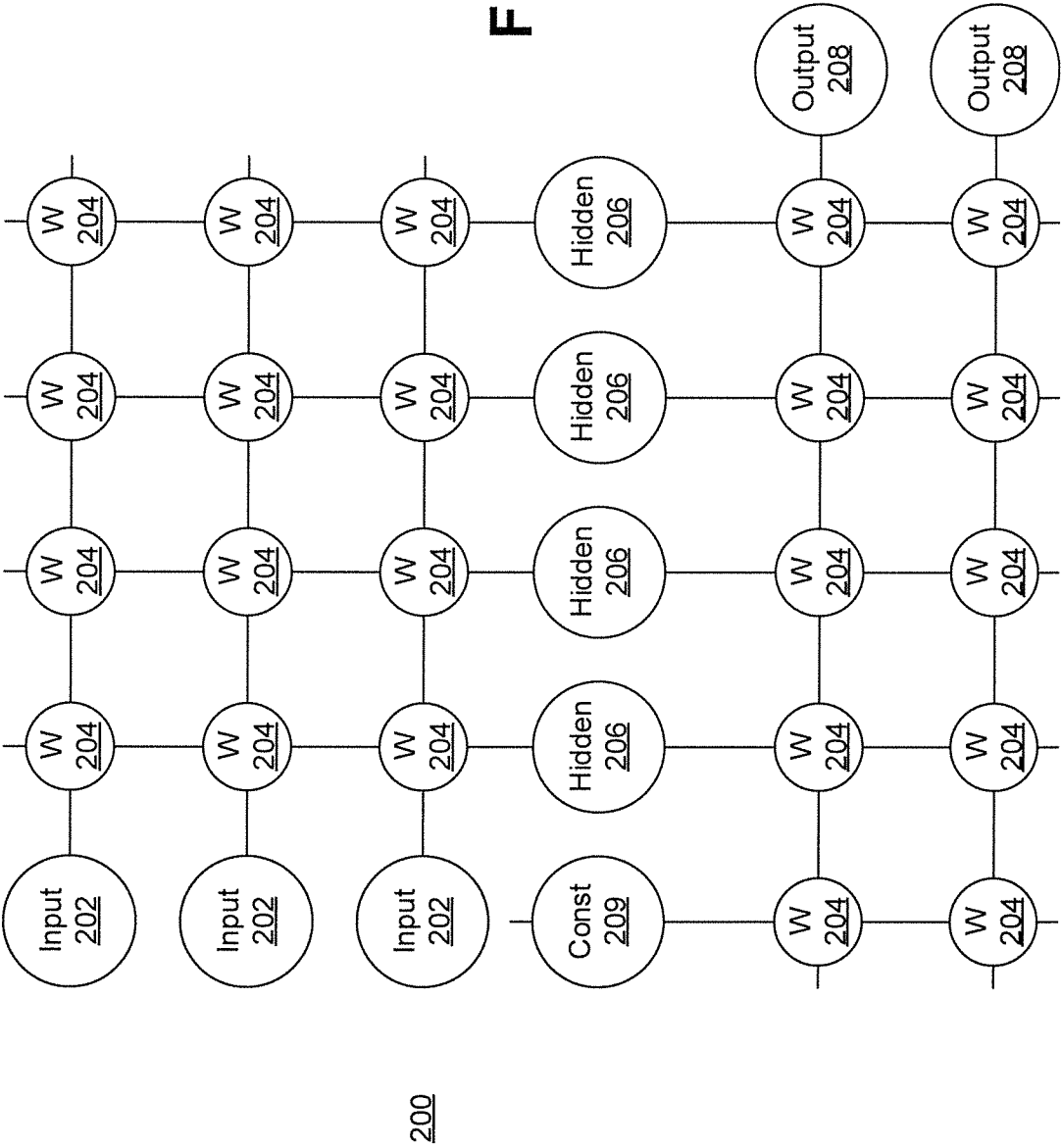


FIG. 1



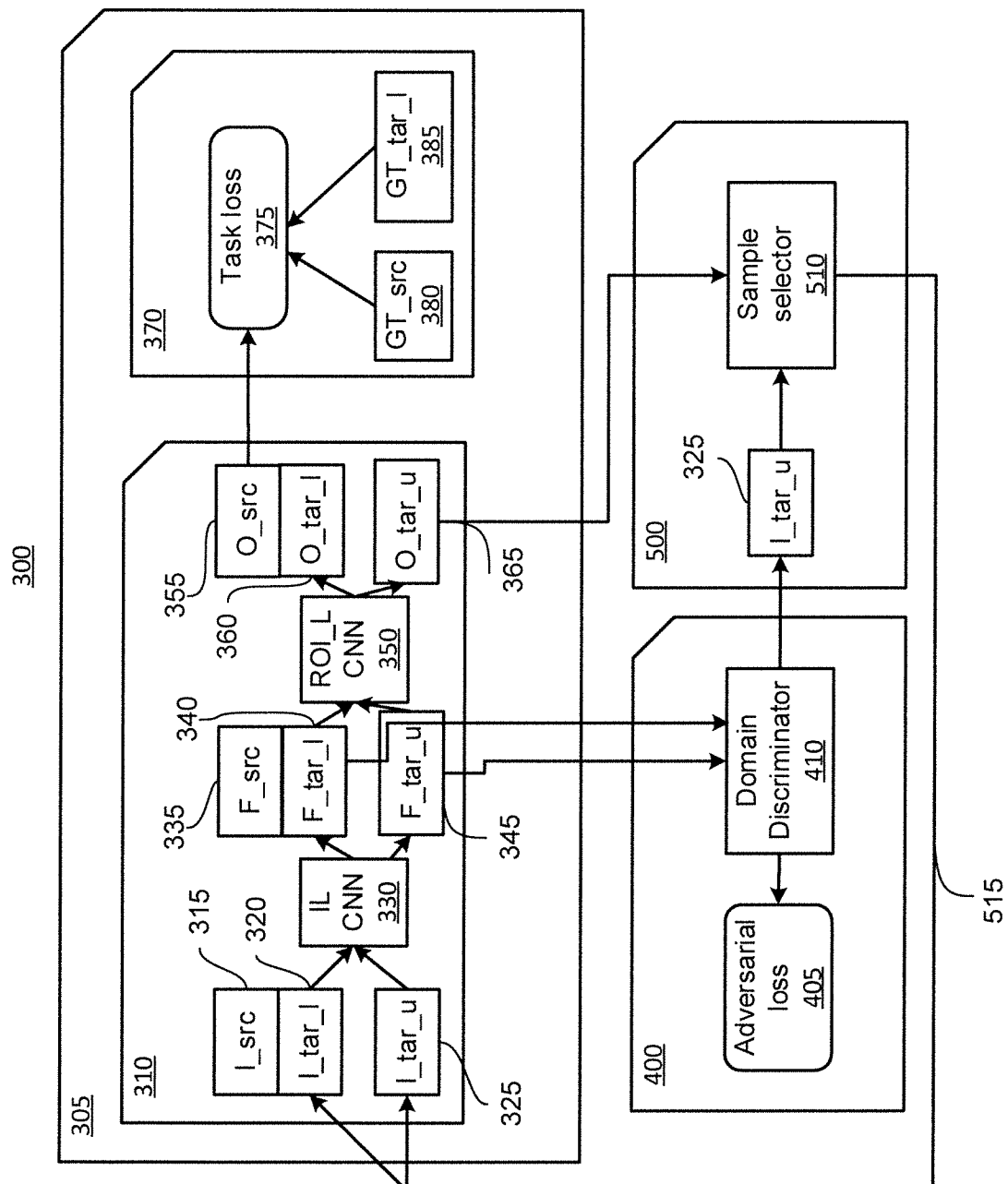


FIG. 3

600

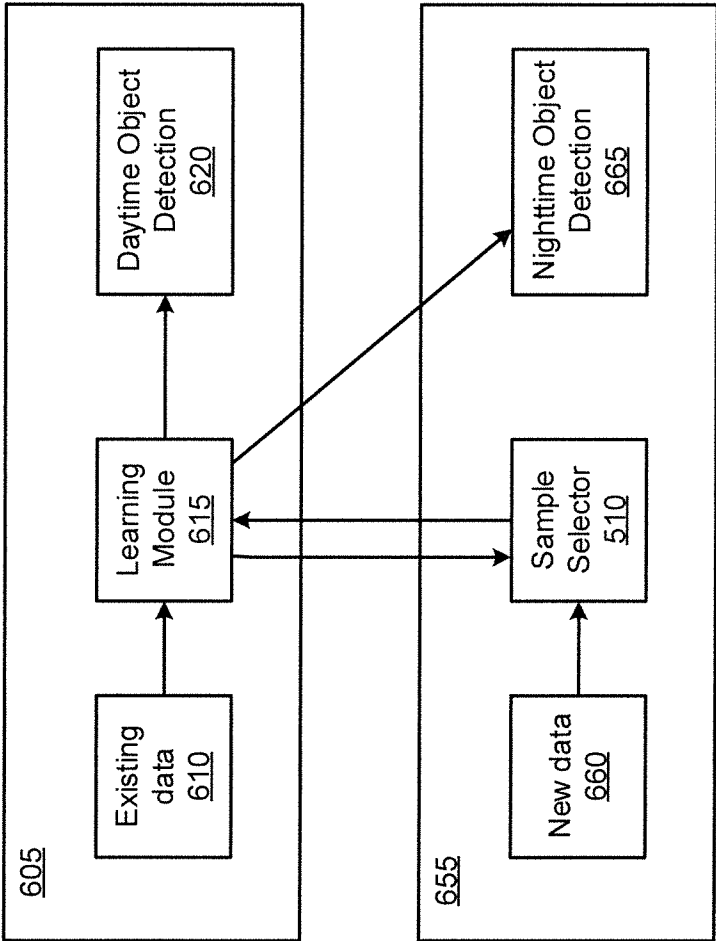


FIG. 4

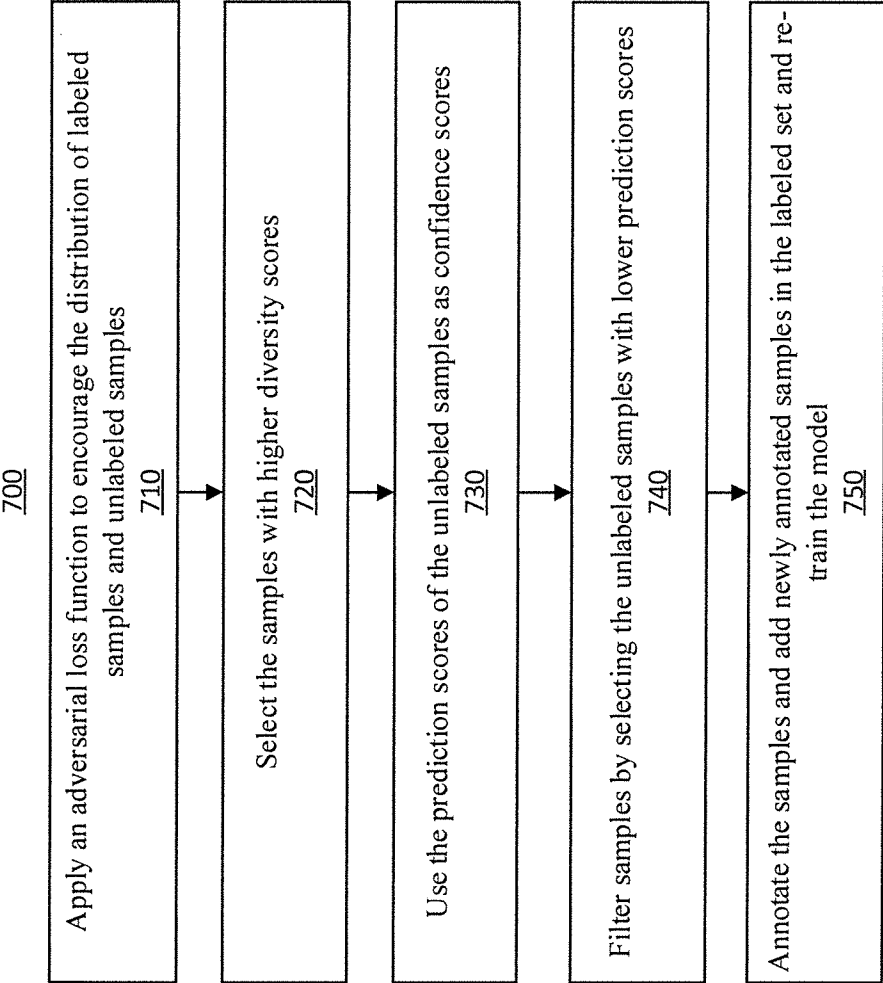


FIG. 5

800

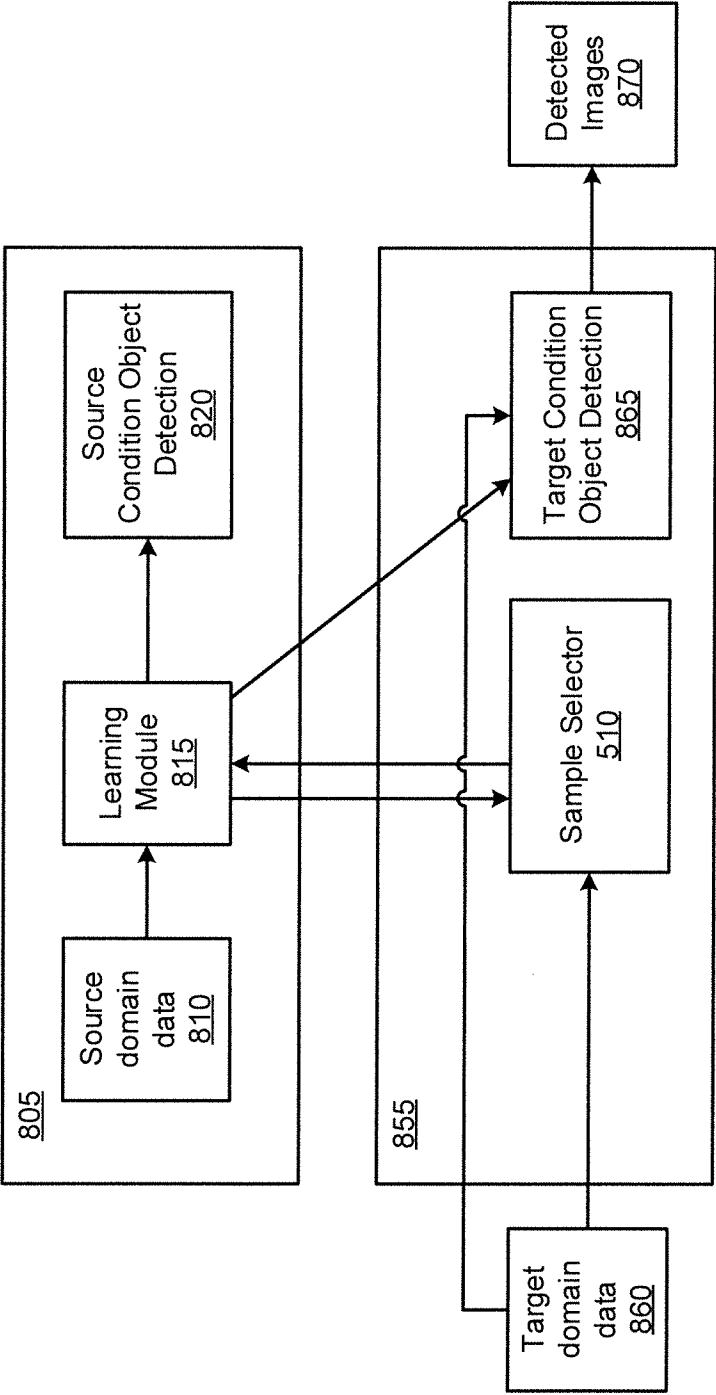


FIG. 6

A. CLASSIFICATION OF SUBJECT MATTER**G06N 3/08(2006.01)i, G06N 3/04(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)

G06N 3/08; G06F 16/00; G06F 17/30; G06K 9/46; G06K 9/62; G06Q 30/06; G06N 3/04

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & keywords: domain adaptation, domain discriminator, active learning, adversarial learning, target domain, source domain, unlabeled samples**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YUE DONG et al., 'Adversarial Active Learning for Sequence Labeling and Generation', Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), 19 July 2018 [retrieved on 2019.12.05]. Retrieved from < https://www.ijcai.org/proceedings/2018/0558.pdf > See page 4014; and figure 1.	1-20
A	US 2018-0211303 A1 (WAL-MART STORES, INC.) 26 July 2018 See paragraphs [0051]-[0059], [0070]-[0072].	1-20
A	KR 10-2018-0097892 A (WISEITECH CO., LTD.) 03 September 2018 See paragraphs [0029]-[0034]; and claim 1.	1-20
A	WO 2018-020277 A1 (GOOGLE LLC) 01 February 2018 See claims 1, 6-8.	1-20
PX	JONG-CHYI SU et al., 'Active Adversarial Domain Adaptation', arXiv:1904.07848v1, 16 April 2019 [retrieved on 2019.12.05]. Retrieved from < https://arxiv.org/abs/1904.07848 > See whole pages.	1-20



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

"D" document cited by the applicant in the international application

"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

17 December 2019 (17.12.2019)

Date of mailing of the international search report

17 December 2019 (17.12.2019)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea



Facsimile No. +82-42-481-8578

Authorized officer

KWON, Sung Ho

Telephone No. +82-42-481-3547



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2019/048722

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018-0211303 A1	26/07/2018	None	
KR 10-2018-0097892 A	03/09/2018	WO 2018-155816 A1	30/08/2018
WO 2018-020277 A1	01/02/2018	CN 109643383 A	16/04/2019
		EP 3475879 A1	01/05/2019
		JP 2019-523504 A	22/08/2019
		KR 10-2019-0031318 A	25/03/2019
		US 2019-0180136 A1	13/06/2019