

METHOD AND DEVICE FOR RE-IDENTIFYING AN OBJECT, TERMINAL AND STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATION

[0001] The disclosure is based upon and claims priority to Chinese Patent Application No. 202010499288.7 filed on June 04, 2020, the disclosure of which is hereby incorporated by reference in its entirety.

TECHNICAL DOMAIN

[0002] The disclosure relates to the technical domain of picture processing, and particularly to an method and device for re-identifying an object, a storage medium and a computer device.

BACKGROUND

[0003] In recent years, domain adaptation strategies have been widely applied to identify, classify, detect pictures and the like in the domain of Artificial Intelligence (AI). Application is, for example, re-identification (ID) of objects (such as pedestrians and vehicles).

[0004] In related art, pseudo-labeling is generally used to implement cross-domain object re-ID. Namely, corresponding true labels are added to source-domain picture data, a network is pre-trained with the source-domain picture data, then target-domain picture data is clustered using the pre-trained network to generate pseudo labels, and finally, the network is optimized using the picture data with the pseudo labels to obtain a final network.

[0005] In related art, when a network is optimized, only picture data with pseudo labels in a target domain is used, and outliers not included in clusters are discarded. However, the outliers may be difficult but valuable sample picture data, so the clustering

performance of the network is restricted, and furthermore, some impacts may be brought to a clustering result of the network.

SUMMARY

[0006] The disclosure provides a method and device for re-identifying an object, a storage medium and a computer device.

[0007] The disclosure provides a method for re-identifying an object, which may include that: a pre-trained re-ID network is acquired; a picture to be identified is acquired; and re-ID is performed on the picture to be identified through the re-ID network to obtain a re-ID result of a target object in the picture to be identified. Training picture data for the re-ID network may at least include first cluster picture data and non-cluster instance picture data, the first cluster picture data and the non-cluster instance picture data may be obtained by performing clustering on a first picture dataset through an initial network corresponding to the re-ID network, and picture data in the first picture dataset may exclude true cluster labels.

[0008] Accordingly, in embodiments of the disclosure, network training is performed in combination with outliers not included in clusters to improve the clustering performance of the re-ID network and further improve the accuracy of the re-ID result, obtained through the method for re-identifying an object of the disclosure, of the target object.

[0009] In an embodiment, the training picture data for the re-ID network may further include a second picture dataset, and second cluster picture data in the second picture dataset may include a true cluster label; and a picture data domain of the second picture dataset may be different from a picture data domain of the first picture dataset.

[0010] Accordingly, in the embodiment of the disclosure, the first cluster picture data excluding true cluster labels, the non-cluster instance picture data and the second cluster picture data including the true cluster label may be supervised to improve the clustering performance of the re-ID network and further improve the accuracy of the re-ID result, obtained through the method for re-identifying an object of the disclosure, of the target

object.

[0011] In an embodiment, before the pre-trained re-ID network is acquired, the method may further include that: the initial network is acquired; the training picture data is acquired; and the initial network is trained through the training picture data to obtain the re-ID network.

[0012] Accordingly, in the embodiment of the disclosure, the initial network may be trained through the acquired training picture data to obtain the re-ID network, so that the picture classification and object ID capabilities of the re-ID network may be improved.

[0013] In an embodiment, the operation that the training picture data is acquired may include that: an initial clustering result obtained by performing clustering on the first picture dataset through the initial network is acquired; and re-clustering is performed on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data.

[0014] Accordingly, processing of target-domain picture data in the embodiment of the disclosure can be understood as a self-spaced contrastive learning strategy. That is, according to a principle of "from simple to difficult", the most trusted cluster is obtained, and then other trusted clusters are gradually obtained by clustering to improve the quality of a learning target. Therefore, errors can be reduced by increasing the number of the trusted clusters.

[0015] In an embodiment, the initial clustering result may include initial cluster picture data. The operation that re-clustering is performed on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data may include that: a picture data volume of a first present cluster in the initial cluster picture data is reduced based on a picture feature distance to obtain a second present cluster; a density index of the second present cluster is determined, the density index being a ratio of a picture data volume of the second present cluster to the picture data volume of the first present cluster; when the density index reaches a first preset threshold, the first present cluster is replaced with the second present cluster to obtain the first cluster picture data; and clipped picture data is classified as non-cluster instance picture data.

[0016] Accordingly, re-clustering may be performed by evaluating densities of

clusters to gradually increase the number of the trusted clusters, thereby improving the quality of the learning target. Therefore, errors can be reduced by increasing the number of the trusted clusters.

[0017] In an embodiment, the initial clustering result may further include initial non-cluster picture data. The operation that re-clustering is performed on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data may include that: picture data of another cluster and/or picture data in the initial non-cluster picture data are/is added to a third present cluster of the initial cluster picture data based on the picture feature distance to obtain a fourth present cluster, the another cluster being different from the third present cluster in the initial cluster picture data; an independence index of the fourth present cluster is determined, the independence index being a ratio of a picture data volume of the third present cluster to a picture data volume of the fourth present cluster; when the independence index reaches the first preset threshold, the third present cluster is replaced with the fourth present cluster to obtain the first cluster picture data; and when the added picture data includes the picture data of the another cluster, the another cluster is canceled, and/or, when the added picture data includes the picture data in the initial non-cluster picture data, the added picture data is not classified as non-cluster instance picture data.

[0018] Accordingly, re-clustering may be performed by evaluating the independence of clusters, so that the ID rate of feature representations can be increased gradually, and more non-cluster data may be added to new clusters to gradually increase the number of the trusted clusters to improve the quality of the learning target. Therefore, errors can be reduced by increasing the number of the trusted clusters.

[0019] In an embodiment, the operation that the initial network is trained through the training picture data to obtain the re-ID network may include that: a picture data center is determined based on the training picture data; a contrastive loss is determined based on the training picture data and the picture data center, and parameter optimization is performed on the initial network based on the contrastive loss to obtain an optimal network; the non-cluster instance picture data in the training picture data is clustered through the optimal network, and the first cluster picture data and the non-cluster instance picture data is updated based on a clustering result to obtain new training picture data; and a new picture data center is determined based on the new training picture data,

and a step that a new contrastive loss is determined based on the new training picture data and the new picture data center is repeated until training is completed, to obtain the re-ID network.

[0020] Accordingly, in the embodiment of the disclosure, the network may be dynamically optimized, the training data may be updated, and the picture data center may be updated, so that the training performance of the re-ID network can be improved, and the accuracy of the re-ID result, obtained through the method for re-identifying an object of the disclosure, of the target object can be further improved.

[0021] In an embodiment, the picture data center may include a first cluster center corresponding to the first cluster picture data and an instance center corresponding to the non-cluster instance picture data; or, the picture data center may include a first cluster center corresponding to the first cluster picture data, an instance center corresponding to the non-cluster instance picture data and a second cluster center corresponding to the second cluster picture data.

[0022] Accordingly, network training may be performed by unsupervised learning, and the second cluster picture data may also be introduced for training by semi-supervised learning, so that the flexibility and diversity of network training are improved.

[0023] In an embodiment, the re-ID network may include a Residual Network (ResNet).

[0024] Accordingly, the ResNet is a network consisting of residual blocks, and skip connection is used for the residual blocks in the network, so that gradient vanishment and gradient explosion may be avoided, the ResNet is endowed with the characteristic of easiness for optimization, and the picture classification and object ID performance may also be improved.

[0025] The disclosure provides a device for re-identifying an object, which may include: a network acquisition module, configured to acquire a pre-trained re-ID network; a picture acquisition module, configured to acquire a picture to be identified; and a Re-ID module, configured to perform re-ID on the picture to be identified through the re-ID network to obtain a re-ID result of a target object in the picture to be identified. Training picture data for the re-ID network may at least include first cluster picture data and non-

cluster instance picture data, the first cluster picture data and the non-cluster instance picture data may be obtained by performing clustering on a first picture dataset through an initial network corresponding to the re-ID network, and picture data in the first picture dataset may exclude true cluster labels.

[0026] The disclosure provides a computer device, which may include memory, a processor and a computer program stored in the memory and capable of running in the processor, the processor executing the program to implement the above method for re-identifying an object.

[0027] The disclosure provides a computer-readable storage medium having stored therein a computer-executable instruction which, when executed by a processor, is configured to implement the method for re-identifying an object.

[0028] The embodiments of the disclosure provide a computer program product including a non-transitory computer-readable storage medium storing a computer program which is operable to enable a computer to execute part or all of the steps described in the above method for re-identifying an object of the embodiments of the disclosure. The computer program product may be a software installer.

[0029] It should be understood that the above general description and the following detailed description are only exemplary and explanatory and not intended to limit the embodiments of the disclosure.

[0030] According to the following detailed descriptions made to exemplary embodiments with reference to the drawings, other features and aspects of the disclosure may become clear.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments consistent with the disclosure and, together with the specification, serve to explain the principle of the disclosure.

[0032] FIG. 1 is a schematic diagram of obtaining a re-ID network by network

training according to an embodiment of the disclosure.

[0033] FIG. 2 is a schematic diagram of processing target-domain picture data according to an embodiment of the disclosure.

[0034] FIG. 3 is a schematic diagram of performing re-clustering on an initial clustering result to obtain first cluster picture data and non-cluster instance picture data according to an embodiment of the disclosure.

[0035] FIG. 4 is an example diagram of calculating a density index according to an embodiment of the disclosure.

[0036] FIG. 5 is a schematic diagram of performing re-clustering on an initial clustering result to obtain first cluster picture data and non-cluster instance picture data according to an embodiment of the disclosure.

[0037] FIG. 6 is an example diagram of calculating an independence index according to an embodiment of the disclosure.

[0038] FIG. 7 is a schematic diagram of training an initial network through training picture data to obtain a re-ID network according to an embodiment of the disclosure.

[0039] FIG. 8 is a schematic diagram of re-identifying an object through a re-ID network according to an embodiment of the disclosure.

[0040] FIG. 9 is a schematic diagram of a method for training a re-ID network according to an embodiment of the disclosure.

[0041] FIG. 10a is a schematic diagram of a re-clustering method according to an embodiment of the disclosure.

[0042] FIG. 10b is a schematic diagram of another re-clustering method according to an embodiment of the disclosure.

[0043] FIG. 11 is a schematic diagram of a device for training a re-ID network according to an embodiment of the disclosure.

[0044] FIG. 12 is a schematic diagram of a device for re-identifying an object

according to an embodiment of the disclosure.

[0045] Specific embodiments of the disclosure have been shown in the drawings, and more detailed descriptions will be made hereinafter. These drawings and text descriptions are not for limiting the scope of the disclosure in any manner but for explaining the concept of the disclosure to those skilled in the art with reference to the specific embodiments.

DETAILED DESCRIPTION

[0046] In order to make the objectives, technical solutions and advantages of the embodiments of the disclosure clearer, the technical solutions in the embodiments of the disclosure will be clearly and completely described below in combination with the drawings in the embodiments of the disclosure. It is apparent that the described embodiments are not all but part of embodiments of the disclosure. All other embodiments obtained by those of ordinary skill in the art based on the embodiments in the disclosure without creative work shall fall within the scope of protection of the disclosure.

[0047] The terms used in the embodiments of the disclosure are only for a purpose of describing specific embodiments and not intended to limit the disclosure. Singular forms "a/an" and "the" used in the embodiments of the disclosure are also intended to include plural forms unless other meanings are clearly expressed in the context.

[0048] It is to be understood that term "and/or" used in the disclosure is only an association relationship describing associated objects and represents that three relationships may exist. For example, A and/or B may represent three conditions: independent existence of A, existence of both A and B and independent existence of B. In addition, character "/" in the disclosure usually represents that previous and next associated objects form an "or" relationship.

[0049] Based on the context, terms "if" and "in case of" used here may be explained as "while" or "when" or "responsive to determining" or "responsive to detecting". Similarly, based on the context, phrase "if determining" or "if detecting (stated condition or event)"

may be explained as "when determining" or "responsive to determining" or "when detecting (stated condition or event)" or "responsive to detecting (stated condition or event)".

[0050] It is also to be noted that terms "include" and "contain" or any other variant thereof is intended to cover nonexclusive inclusions, so that a product or system including a series of elements not only includes those elements but also includes other elements which are not clearly listed or further includes elements intrinsic to the product or the system. Under the condition of no more limitations, an element defined by the statement "including a/an....." does not exclude existence of the same other elements in a product or system including the element.

[0051] AI is a theory, method, technology and application system of simulating, extending and developing human intelligence using digital computers or machines controlled by digital computers to sense environments, acquire knowledge and obtain optimal results using the knowledge. In other words, AI is an integrated technology of computer science, aiming at understanding the essence of intelligence and generating a novel intelligent machine capable of responding in a manner similar to human intelligence. AI can research design principles and implementation methods of various intelligent machines to endow the machines with sensing, inference and decision-making functions.

[0052] The AI technology, as a comprehensive subject covering a wide range of domains, includes both hardware and software technologies. Basic AI technologies include sensors, dedicated AI chips, cloud computing, distributed storage, big data processing, operating/interactive systems, mechatronics, and other technologies. AI software technologies mainly have some major directions such as Computer Vision (CV) and Machine Learning (ML)/deep learning.

[0053] In some embodiments of the disclosure, CV, as a science researching how to make machines "see", refers to implementing machine vision such as identification, tracking, and measurement on targets using video cameras and computers instead of human eyes and performing graphic processing through computers to obtain pictures more suitable for human eyes to view or to be sent to instruments for detection. As a science subject, CV researches related theories and technologies for establishing AI

systems capable of acquiring information from pictures or multidimensional data. CV usually includes picture processing, picture identification, semantic picture understanding, picture retrieval, Optical Character Recognition (OCR), video processing, semantic video understanding, video content/behavior ID, Three Dimensional (3D) object reconstruction, 3D technology, virtual reality, augmented reality, simultaneous localization and mapping, and other technologies, and also includes common biometric feature identification technologies such as face identification and fingerprint identification.

[0054] Machine Learning (ML), as an interdisciplinary subject involving many disciplines such as probability theory, statistics, approximation theory, convexity analysis, and algorithm complexity theory, specifically researches how computers simulate or learn behaviors of humans to acquire new knowledge or skills and reorganize existing knowledge structures to constantly improve their own performance. ML is the core of AI as well as the basic solution of making computers intelligent, and can be applied to various domains of AI extensively. ML and deep learning usually include technologies such as artificial neural network, belief network, reinforcement learning, inductive learning, and learning from demonstration.

[0055] Target re-ID, as an important issue in the domains of CV and security monitoring, requires retrieval of a picture of a corresponding target from a dataset, and the target may be a pedestrian, a vehicle, etc. However, when a trained network is directly applied to different monitoring scenarios, the network may present inevitable performance reduction caused by differences between picture domains, such as shooting environments, light, backgrounds, and shooting equipment. In addition, it is impractical to label different training data for network training in each monitoring scenario, and this is because labeling costs a lot of manpower and time.

[0056] At present, a pseudo-labeling-based method is a common method for target re-ID based on domain adaptation (a kind of transfer learning). The method aims to keep performing clustering in an unlabeled target domain to generate pseudo labels for self-training, and the most advanced performance may be achieved. However, some outliers, i.e., edge samples that cannot be divided into any class, may be generated in a clustering process. In a related method, for ensuring the clustering quality, these outliers may be directly discarded and not added to a training set, namely in a self-training process of a network, only picture data with pseudo labels in the target domain are used, and outliers

not included in clusters are discarded. However, the outliers may be difficult but valuable sample picture data, so the clustering performance of the network is restricted, and furthermore, some impacts may be brought to a clustering result of the network.

[0057] Based on this, the disclosure provides a method for re-identifying an object. A re-ID network used in the method may be obtained by training at least first cluster picture data and non-cluster instance picture data. Therefore, in the disclosure, network training may be performed in combination with outliers not included in clusters to improve the clustering performance of the re-ID network and further improve the accuracy of re-ID result of a target object obtained through the re-ID method of the disclosure.

[0058] The method for re-identifying an object in the embodiments of the disclosure may be divided into two parts, including a network training part and a network application part. The network training part involves the technical domain of ML, and in the network training part, an initial network is trained through the ML technology to obtain the trained re-ID network. In the network application part, a re-ID result of a target object in a picture to be identified may be obtained using the re-ID network trained in the network training part.

[0059] For convenient understanding, the network training part in the solution of the disclosure will be explained and described at first.

[0060] It can be understood that steps of the network training part in the disclosure may be implemented by a terminal or a server.

[0061] FIG. 1 is a schematic diagram of obtaining a re-ID network by network training according to an embodiment of the disclosure. As shown in FIG. 1, the processing flow includes the following steps.

[0062] In S100, an initial network is acquired.

[0063] In S200, training picture data is acquired.

[0064] In S300, the initial network is trained through the training picture data to obtain a re-ID network.

[0065] The initial network is a network initially to be trained, and the initial network

has an object re-ID capability.

[0066] The initial network may be a ResNet, etc. The ResNet is a network consisting of residual blocks, and skip connection is used for residual blocks in the network, so that gradient vanishment and gradient explosion may be avoided, the ResNet is endowed with the characteristic of easiness for optimization, and the picture classification and object ID performance is also be improved.

[0067] In some embodiments, unsupervised learning may be applied in a network training method. Unsupervised learning refers to a process of performing network training only with unlabeled picture data in a target domain. The target domain may be a first monitoring scenario.

[0068] Under the condition of performing network training by unsupervised learning, the training picture data for the re-ID network includes first cluster picture data and non-cluster instance picture data. The first cluster picture data and the non-cluster instance picture data may be obtained by performing clustering on a first picture dataset through the initial network corresponding to the re-ID network. Picture data in the first picture dataset exclude true cluster labels, and the first picture dataset corresponds to the picture data of the target domain.

[0069] In some embodiments, semi-supervised learning may be applied in the network training method. Semi-supervised learning refers to a process of performing network training with both labeled picture data in a source domain and unlabeled picture data in the target domain. The source domain may be a second monitoring scenario. The labeled picture data in the source domain have ground-truth labels, and ground-truth may be manually labeled, and ground-truth may provide valuable supervision in the network training process.

[0070] Under the condition of performing network training by semi-supervised learning, the training picture data for the re-ID network at least include the first cluster picture data, the non-cluster instance picture data and a second picture dataset.

[0071] The first cluster picture data and the non-cluster instance picture data may be obtained by performing clustering on the first picture dataset through the initial network corresponding to the re-ID network. The picture data in the first picture dataset exclude

true cluster labels, and the first picture dataset corresponds to the picture data of the target domain.

[0072] Second cluster picture data in the second picture dataset includes a true cluster label, and the second picture dataset corresponds to the picture data of the source domain. A picture data domain of the second picture dataset is different from a picture data domain of the first picture dataset.

[0073] In an embodiment, under the condition of performing network training by semi-supervised learning, the step that the training picture data is acquired may include steps that labeled source-domain picture data is acquired, unlabeled target-domain picture data is acquired, and the target-domain picture data is processed.

[0074] When the source-domain picture data is acquired, the labeled picture data may be directly acquired.

[0075] In some embodiments, under the condition of performing network training by unsupervised learning, the step that the training picture data is acquired may include the steps that the unlabeled target-domain picture data is acquired and the target-domain picture data is processed.

[0076] FIG. 2 is a schematic diagram of processing target-domain picture data. As shown in FIG. 2, the processing flow includes the following steps.

[0077] In S220, an initial clustering result obtained by performing clustering on a first picture dataset through the initial network is acquired.

[0078] In S240, re-clustering is performed on the initial clustering result to obtain first cluster picture data and non-cluster instance picture data.

[0079] The first picture dataset corresponds to the target-domain picture data. After the unlabeled target-domain picture data is acquired, initial clustering is performed on the first picture dataset through the initial network at first to obtain the initial clustering result corresponding to the first picture dataset, and then re-clustering is performed on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data.

- [0080] The above flow of processing the target-domain picture data can be understood as a self-spaced contrastive learning strategy. That is, according to a principle of "from simple to difficult", the most trusted cluster is obtained, and then other trusted clusters may be gradually obtained by clustering to improve the quality of a learning target. Therefore, errors can be reduced by increasing the number of the trusted clusters.
- [0081] In an embodiment, there is provided a clustering credibility evaluation criterion. According to the criterion, re-clustering may be performed by evaluating densities of clusters, thereby increasing the number of trusted clusters.
- [0082] In the embodiment, the initial clustering result may include initial cluster picture data.
- [0083] FIG. 3 is a schematic diagram of performing re-clustering on an initial clustering result to obtain first cluster picture data and non-cluster instance picture data according to an embodiment of the disclosure. As shown in FIG. 3, the processing flow includes the following steps.
- [0084] In S242A, a picture data volume of a first present cluster in the initial cluster picture data is reduced based on a picture feature distance to obtain a second present cluster.
- [0085] In S244A, a density index of the second present cluster is determined, the density index being a ratio of a picture data volume of the second present cluster to the picture data volume of the first present cluster.
- [0086] In S246A, when the density index reaches a first preset threshold, the first present cluster is replaced with the second present cluster to obtain the first cluster picture data.
- [0087] In S248A, clipped picture data is classified as non-cluster instance picture data.
- [0088] In the embodiment, a clustering standard is proposed to perform re-clustering to verify whether the densities of the clusters meet a preset requirement or not.
- [0089] For all the picture data divided into the same cluster, it can be understood that a picture feature distance of the picture data meets the clustering standard, namely

$\Delta d \leq d^1$, where Δd is the picture feature distance, and d^1 is a distance corresponding to the clustering standard.

[0090] After the clustering standard is raised (the distance corresponding to the clustering standard is reduced), for example, the clustering standard is changed to d^2 , $d^2 < d^1$, a picture feature distance of part of picture data may be greater than the clustering standard, namely $\Delta d > d^2$. In such a case, picture data corresponding to $\Delta d \leq d^2$ is retained based on the picture feature distance, and picture data corresponding to $\Delta d > d^2$ is removed from the first present cluster to reduce the picture data volume of the first present cluster to obtain a second present cluster.

[0091] After the second present cluster is obtained, the density index of the second present cluster is calculated, the density index being configured to evaluate a density of the cluster. The density index may be calculated through the following formula: $P = n2 / n1$, where P is the density index, n1 is the picture data volume of the first present cluster, and n2 is the picture data volume of the second present cluster.

[0092] FIG. 4 is an example diagram of calculating a density index. As shown in FIG. 4, the dots represent the picture data, the black dots represent the retained picture data, the white dots represent the removed picture data, the region circled by the solid line represents the first present cluster clu1, and the region circled by the dotted line represents the second present cluster clu2. It can be seen from FIG. 4 that the picture data volume of the first present cluster clu1 is 7 and the picture data volume of the second present cluster clu2 is 5. In such a case, the density index P of the second present cluster clu2 is $P = n2 / n1 = 5 / 7$.

[0093] After the density index P is calculated, the density index P is compared with the corresponding first preset threshold ΔP , and whether to retain the new cluster (i.e., the second present cluster) is determined according to a comparison result.

[0094] $P \geq \Delta P$ indicates that the density index P of the second present cluster clu2 meets a preset density requirement. In such a case, the first present cluster is canceled, the second present cluster is retained, and the first cluster picture data is updated with the second present cluster. Meanwhile, the clipped (removed) picture data in the cluster are

classified as non-cluster instance picture data. For example, referring to FIG. 4, when P is $5/7$ and ΔP is 0.5 , $P > \Delta P$, and in such a case, the first present cluster is replaced with the second present cluster to update the first cluster picture data.

[0095] $P < \Delta P$ indicates that the density index P of the second present cluster $clu2$ does not meet the preset density requirement. In such a case, the second present cluster is canceled, and the first present cluster is retained.

[0096] In the embodiment, re-clustering may be performed by evaluating densities of clusters to gradually increase the number of the trusted clusters, thereby improving the quality of the learning target. Therefore, errors can be reduced by increasing the number of the trusted clusters.

[0097] In an embodiment, there is provided a clustering credibility evaluation criterion. According to the criterion, the independence of clusters is evaluated to perform re-clustering on the initial clustering result, thereby increasing the number of the trusted clusters.

[0098] In the embodiment, the initial clustering result includes the initial cluster picture data and initial non-cluster picture data.

[0099] FIG. 5 is a schematic diagram of performing re-clustering on an initial clustering result to obtain first cluster picture data and non-cluster instance picture data according to an embodiment of the disclosure. As shown in FIG. 5, the processing flow includes the following steps.

[00100] In S242B, picture data of another cluster and/or picture data in the initial non-cluster picture data are/is added to a third present cluster of the initial cluster picture data based on the picture feature distance to obtain a fourth present cluster, the another cluster being different from the third present cluster in the initial cluster picture data.

[00101] In S244B, an independence index of the fourth present cluster is determined, the independence index being a ratio of a picture data volume of the third present cluster to a picture data volume of the fourth present cluster.

[00102] In S246B, when the independence index reaches the first preset threshold, the

third present cluster is replaced with the fourth present cluster to obtain the first cluster picture data.

[00103] In S248B, when the added picture data includes the picture data of the another cluster, the another cluster is canceled, and/or, when the added picture data includes the picture data in the initial non-cluster picture data, the added picture data is not classified as non-cluster instance picture data.

[00104] In the embodiment, the clustering standard may be lowered to perform re-clustering to verify whether the independence of the clusters meets a preset requirement or not.

[00105] For all the picture data divided into the same cluster, it can be understood that a picture feature distance of the picture data meets the clustering standard, namely $\Delta d \leq d1$, where Δd is the picture feature distance, and $d1$ is the distance corresponding to the clustering standard.

[00106] After the clustering standard is lowered (the distance corresponding to the clustering standard is increased), for example, the clustering standard is changed to $d3$, $d3 > d1$, a picture feature distance of picture data not in a present cluster (for example, picture data in another cluster and/or picture data in the initial non-cluster picture data) may reach the clustering standard, namely $\Delta d' \leq d3$, where $\Delta d'$ is the picture feature distance of the picture data not in the present cluster.

[00107] In such a case, the picture data not in the present cluster corresponding to $\Delta d' \leq d3$ is added to the third present cluster based on the picture feature distance to increase the picture data volume of the third present cluster to obtain the fourth present cluster.

[00108] It can be understood that the added picture data may include satisfactory picture data in the present cluster only, or may include satisfactory picture data in the initial non-cluster picture data only, or may include both the satisfactory picture data in another cluster and the satisfactory picture data in the initial non-cluster picture data.

[00109] After the fourth present cluster is obtained, the independence index of the fourth present cluster is calculated, the independence index being configured to evaluate

the independence of the cluster. The independence index may be calculated through the following formula: $Q = n3 / n4$, where Q is the independence index, $n3$ is the picture data volume of the third present cluster, and $n4$ is the picture data volume of the fourth present cluster.

[00110] FIG. 6 is an example diagram of calculating an independence index. As shown in FIG. 6, the regions circled by the solid lines represent clusters obtained before re-clustering, i.e., clusters in the initial cluster picture data, including the third present cluster $clu3$ and another cluster $clui$; the dots represent the picture data; the black dots represent the picture data in the initial cluster picture data; the white dots represent the picture data in the initial non-cluster picture data; and the region circled by the dotted line represents the fourth present cluster $clu4$. It can be seen from FIG. 6 that the picture data volume of the third present cluster $clu3$ is 2 and the picture data volume of the fourth present cluster $clu4$ is 7. In such a case, the independence index Q of the fourth present cluster $clu4$ is $Q = n3 / n4 = 2 / 7$.

[00111] After the independence index Q is calculated, the independence index Q is compared with a corresponding second preset threshold ΔQ , and it is determined whether to retain the new cluster (i.e., the fourth present cluster) or not according to a comparison result.

[00112] $Q \geq \Delta Q$ indicates that the independence index Q of the fourth present cluster $clu4$ meets a preset independence requirement. In such a case, the third present cluster is canceled, the fourth present cluster is retained, and the first cluster picture data is updated with the fourth present cluster.

[00113] When the added picture data includes the picture data of another cluster, the another cluster is canceled. For example, when the independence index Q of the fourth present cluster $clu4$ meets the preset independence requirement, another cluster $clui$ (i is an integer representing the serial number of the cluster) is canceled.

[00114] When the added picture includes the picture data in the initial non-cluster picture data, the added picture data is not classified as non-cluster instance picture data.

[00115] $Q < \Delta Q$ indicates that the independence index Q of the fourth present cluster

clu4 does not meet the preset independence requirement. In such a case, the fourth present cluster is canceled, and the third present cluster is retained.

[00116] When the added picture data includes the picture data of another cluster, the another cluster is retained. For example, when the independence index Q of the fourth present cluster clu4 does not reach the preset independence requirement, the another cluster clui is retained.

[00117] When the added picture includes the picture data in the initial non-cluster picture data, the added picture data is classified as non-cluster instance picture data.

[00118] For example, referring to FIG. 6, when Q is $2/7$ and ΔQ is 0.5 , $Q < \Delta P$. In such a case, the fourth present cluster clu3 is canceled, the third present cluster clu3 and another cluster clui are retained, and the added unclustered picture data is classified as non-cluster instance picture data.

[00119] In the embodiment, re-clustering may be performed by evaluating the independence of clusters, so that the ID rate of feature representations may be increased gradually, and more non-cluster data may be added to new clusters to gradually increase the number of the trusted clusters to improve the quality of the learning target. Therefore, errors can be reduced by increasing the number of the trusted clusters.

[00120] In an embodiment, there is provided a clustering credibility evaluation criterion. According to the criterion, re-clustering may be performed on the initial clustering result by evaluating the independence and densities of clusters, thereby increasing the number of the trusted clusters.

[00121] A flow of performing re-clustering on the initial clustering result based on the independence and the densities may refer to the steps of performing re-clustering by evaluating the independence of the cluster and evaluating the density of the cluster respectively in the abovementioned embodiments, and is not elaborated here.

[00122] When the independence and the density are both taken into account for re-clustering, corresponding preset thresholds may be set according to a practical condition. For example, both ΔP and ΔQ are set to be 0.5 .

- [00123] In the embodiment, re-clustering may be performed by evaluating both the independence and densities of clusters to gradually increase the number of the trusted clusters, thereby improving the quality of the learning target. Therefore, errors can be reduced by increasing the number of the trusted clusters.
- [00124] In an embodiment, steps for network training are explained and described.
- [00125] FIG. 7 is a schematic diagram of obtaining an initial network through training picture data to obtain a re-ID network according to an embodiment of the disclosure. As shown in FIG. 7, the processing flow includes the following steps.
- [00126] In S320, a picture data center is determined based on the training picture data.
- [00127] In S340, a contrastive loss is determined based on the training picture data and the picture data center, and parameter optimization is performed on the initial network based on the contrastive loss to obtain an optimal network.
- [00128] In S360, the non-cluster instance picture data in the training picture data is clustered through the optimal network, and the first cluster picture data and the non-cluster instance picture data are updated based on a clustering result to obtain new training picture data.
- [00129] In S380, a new picture data center is determined based on the new training picture data, and a step that a new contrastive loss is determined based on the new training picture data and the new picture data center is repeated until training is completed, to obtain the re-ID network.
- [00130] In some embodiments, under the condition of performing network training by unsupervised learning, the training data includes the first cluster picture data and the non-cluster instance picture data. Correspondingly, the picture data center includes a first cluster center corresponding to the first cluster picture data and an instance center corresponding to the non-cluster instance picture data.
- [00131] In some embodiments, under the condition of performing network training by semi-supervised learning, the training data includes the first cluster picture data, the non-cluster instance picture data and the second cluster picture data. Correspondingly, the picture data center includes the first cluster center corresponding to the first cluster

picture data, the instance center corresponding to the non-cluster instance picture data and a second cluster center corresponding to the second cluster picture data.

[00132] Explanations and descriptions are made with performing network training by semi-supervised learning as an example.

[00133] (1) An initial picture data center is determined based on the acquired training picture data at first.

[00134] When the corresponding first cluster center is determined based on the first cluster picture data, an average feature vector of picture data in a cluster in the first cluster picture data may be taken as the first cluster center corresponding to the cluster. It can be understood as that, when the first cluster picture data includes multiple clusters, there are correspondingly multiple first cluster centers.

[00135] When the corresponding instance center is determined based on the non-cluster instance picture data, for each separate instance in the non-cluster instance picture data, a feature vector corresponding to the separate instance is an instance center of the separate instance. It can be understood as that, when the non-cluster instance picture data includes multiple separate instances, there are correspondingly multiple instance centers.

[00136] When the corresponding second cluster center is determined based on the second cluster picture data, an average feature vector of picture data in a cluster in the second cluster picture data may be taken as the second cluster center corresponding to the cluster. It can be understood as that, when the second cluster picture data includes multiple clusters, there are correspondingly multiple second cluster centers.

[00137] (2) The contrastive loss is determined based on the training picture data and the picture data center, and parameter optimization is performed on the initial network based on the contrastive loss to obtain the optimal network.

[00138] It is defined that X^s represents the second cluster picture data (i.e., the source-domain data) in the second picture dataset, X^t represents the first picture dataset (i.e., the target-domain data), X_c^t represents the first cluster picture data and X_0^t represents the non-cluster instance picture data. In such a case, $X^t = X_c^t \cup X_0^t$.

[00139] For the feature vector $f = f\theta(x)$, $x \in X^S \cup X_c^t \cup X_o^t$, the contrastive loss may be calculated through the following formula (1), and parameter optimization is performed on the initial network based on the contrastive loss to obtain the optimal network:

[00140]
$$L_f = -\log \frac{\exp(\langle f, Z^+ \rangle / \tau)}{\sum_{k=1}^{n^s} \exp(\langle f, w_k \rangle / \tau) + \sum_{k=1}^{n_c^t} \exp(\langle f, c_k \rangle / \tau) + \sum_{k=1}^{n_o^t} \exp(\langle f, v_k \rangle / \tau)}$$

 (1).

[00141] τ is set to be 0.05. $\langle a, b \rangle$ represents an inner product between two feature vectors a and b , and is used to measure a similarity between the feature vectors. n^s represents the number of clusters in the second cluster picture data. n_c^t represents the number of clusters in the first cluster picture data. n_o^t represents the number of separate instances in the non-cluster instance picture data. w_k represents the second cluster center corresponding to the second cluster picture data. c_k represents the first cluster center corresponding to the first cluster picture data. v_k represents the instance center corresponding to the non-cluster instance picture data.

[00142] In addition, Z^+ represents the data center corresponding to the feature vector f . For example, $Z^+ = w_k$ when $f = f\theta(x)$, $x \in X^S$, $Z^+ = c_k$ when $f = f\theta(x)$, $x \in X_c^t$, and $Z^+ = v_k$ when $f = f\theta(x)$, $x \in X_o^t$.

[00143] (3) After the optimal network is obtained, the non-cluster instance picture data is clustered through the optimal network, and the first cluster picture data and the non-cluster instance picture data are updated according to the clustering result.

[00144] In the processing process of the disclosure, hybrid memory may store the first cluster picture data, the non-cluster instance picture data and the second cluster picture data, as well as the first cluster center corresponding to the first cluster picture data, the instance center corresponding to the non-cluster instance picture data and the second cluster center corresponding to the second cluster picture data.

[00145] It can be understood that, during each iteration, feature vectors processed every time participate in updating of the hybrid memory.

[00146] During clustering using the optimal network, a new clustering result may be obtained, so there may be brought an update of the first cluster picture data and the non-cluster instance picture data, namely new training picture data is obtained. After the new training picture data is obtained, the hybrid memory may be updated accordingly.

[00147] (4) After the new training picture data is obtained, the new picture data center is determined based on the new training picture data, namely the picture data center stored in the hybrid memory is updated and regulated.

[00148] It can be understood that the second cluster center may be updated by regulation based on an original center, while the first cluster center and the instance center may be updated by recalculation according to the update of the first cluster picture data and the non-cluster instance picture data.

[00149] Updating of the second cluster center w_k may be implemented through the following formula (2):

$$[00150] \quad w_k \leftarrow m^s w_k + (1 - m^s) \cdot \frac{1}{|\beta_k|} \sum_{f_i^s \in \beta_k} f_i^s \quad (2).$$

[00151] β_k is a feature of the second cluster picture data during present processing, and $m^s \in [0,1]$ is a momentum coefficient for updating the second cluster center. For example, m^s may be set to be 0.2.

[00152] Updating of the first cluster center c_k may be implemented through the following formula (3):

$$[00153] \quad c_k = \frac{1}{|I_k|} \sum_{v_i \in I_k} v_i \quad (3).$$

[00154] I_k is a kth cluster in the first cluster picture data, and $|I_k|$ represents the

number of features in the cluster.

[00155] Updating of the instance center V_k may be implemented through the following formula (4):

[00156]
$$V_i \leftarrow m^t V_i + (1 - m^t) F_i^t \quad (4).$$

[00157] $m^t \in [0,1]$ is a momentum coefficient for updating the instance center. For example, m^t may be set to be 0.2.

[00158] Picture data in the non-cluster instance picture data is given. Under the condition of determining, through the optimal network, that the picture data belong to the kth cluster, the first cluster center C_k is updated using the updating formula for the first cluster center C_k .

[00159] (5) After the hybrid memory is updated, (2) is repeated to perform network iterative training, to obtain the re-ID network when the network converges.

[00160] In an embodiment, under the condition of performing network training by unsupervised learning, the training picture data excludes the second cluster picture data, the picture data center does not include the second cluster center corresponding to the second cluster picture data, and the principle is similar to the principle of performing network training by semi-supervised learning, and is not elaborated here.

[00161] In an embodiment, the network application part in the solution of the disclosure is explained and described.

[00162] It can be understood that method steps of the network application part in the disclosure may be implemented by the terminal or the server, and the execution body of the method steps of the network application part may be the same as or different from the execution body of the method step of the network training part.

[00163] FIG. 8 is a schematic diagram of re-identifying an object through a re-ID network according to an embodiment of the disclosure. As shown in FIG. 8, the processing flow includes the following steps.

- [00164] In S400, the pre-trained re-ID network is acquired.
- [00165] In S500, a picture to be identified is acquired.
- [00166] In S600, re-ID is performed on the picture to be identified through the re-ID network to obtain a re-ID result of a target object in the picture to be identified.
- [00167] The re-ID network may be obtained by training through the method steps of the network training part in each of the above embodiments of the disclosure.
- [00168] When the re-ID network is trained by unsupervised learning, the training picture data for the re-ID network at least include the first cluster picture data and the non-cluster instance picture data. The first cluster picture data and the non-cluster instance picture data may be obtained by performing clustering on the first picture dataset through the initial network corresponding to the re-ID network, and picture data in the first picture dataset exclude true cluster labels.
- [00169] When the re-ID network is trained by semi-supervised learning, the picture data for the re-ID network further includes the second picture dataset. The second cluster picture data in the second picture dataset include true cluster labels, and the picture data domain of the second picture dataset is different from the picture data domain of the first picture dataset.
- [00170] The embodiment provides a method for re-identifying an object. The re-ID network used in the method may be obtained by training at least the first cluster picture data and the non-cluster instance picture data. Therefore, in the disclosure, network training may be performed in combination with outliers not included in clusters to improve the clustering performance of the re-ID network and further improve the accuracy of the re-ID result of the target object obtained through the re-ID method of the disclosure.
- [00171] Target re-ID, as an important issue in the fields of CV and security monitoring, requires retrieval of a picture of a corresponding target from a dataset, and the target may be a pedestrian, a vehicle, etc. However, when a trained model is directly applied to different monitoring scenarios, the model presents inevitable performance reduction caused by differences between domains, such as shooting environments, light,

backgrounds, and shooting equipment. In addition, it is unpractical to label different training data for network training in each monitoring scenario, and this is because labeling costs a lot of manpower and time.

[00172] Unsupervised DA aims to transfer a model trained with labeled data in a source domain to an unlabeled target domain such that the model may learn distinctive features in the target domain to implement target re-ID effectively. The source domain may be monitoring scenario A, and the target domain may be monitoring scenario B. Since targets in the source domain and the target domain have different identities, problems related to unsupervised DA for target re-ID are problems of open sets. The target may be a pedestrian, a vehicle, etc.

[00173] Completely unsupervised re-ID aims to learn distinctive features without any labeled data, namely directly implementing target re-ID in the target domain in an unsupervised manner effectively without the labeled data in the source domain.

[00174] At present, a pseudo-labeling-based method is the most effective method for unsupervised or unsupervised DA-based target re-ID and aims to keep performing clustering in an unlabeled target domain to generate pseudo labels for self-training, to obtain the most advanced performance. However, the method has the following defects that restrict the performance improvement thereof. First, since some exceptionally clustered samples, i.e., edge samples that cannot be divided into any class, may be generated in a clustering process, in an existing method, and for ensuring the clustering quality, these exceptionally clustered samples may be directly discarded and not added to a training set. However, these exceptionally clustered samples may be considered as valuable difficult samples that should be learned. Second, a clustering-based unsupervised DA algorithm usually is adopted to perform pre-training with data of a source domain, then read a trained model, and perform training through pseudo labels generated by clustering and unlabeled target-domain samples for transferring to a target domain. In this algorithm, in a training process in the target domain, valuable source-domain data is discarded and data with true labels in the source domain are wasted, so there is brought a performance loss of the source domain. Third, unsupervised DA-based target re-ID is deficient in ID degree, and unsupervised target re-ID has never been explored. Fourth, a related contrastive learning loss function only considers instance-level supervision.

- [00175] The embodiments of the disclosure provide a self-spaced contrastive learning method for unsupervised target re-ID. A unified contrastive learning framework is provided to simultaneously perform feature learning on all samples in a source domain and a target domain. The framework dynamically updates a hybrid memory module, thereby simultaneously providing source-domain true class-level, target-domain cluster-level and a target-domain non-cluster instance-level supervisions.
- [00176] The embodiments of the disclosure provide a self-spaced contrastive learning strategy and a novel clustering credibility evaluation criterion, to reduce training errors by trusted clustering. According to the strategy, more trusted clusters can be generated gradually to improve feature learning, thereby obtaining more effective features to contribute to clustering.
- [00177] According to the methods in the embodiments of the disclosure, a high ID degree is achieved for unsupervised DA-based pedestrian and vehicle re-ID, and the performance of a source domain may be improved effectively without manual labeling. The methods of the embodiments of the disclosure may be simply extended to unsupervised target re-ID, namely source-domain data and class-level supervision of the source domain in training may be removed to improve the performance remarkably, compared with the related method.
- [00178] The unified contrastive learning framework in the embodiments of the disclosure includes a convolutional neural network-based picture coder and a hybrid memory module. The hybrid memory module may be dynamically updated based on a picture feature output by the picture coder, and may provide source-domain true class-level supervision, target-domain cluster-level supervision and target-domain non-cluster instance-level supervision in real time. Specifically, the hybrid memory module may use a class centroid of the source domain, a cluster centroid of the target domain and a non-cluster instance feature of the target domain for supervision. A coding feature of the source domain is used to directly update the class centroid of the source domain, a coding feature of the target domain is used to update the instance-level feature, and the cluster centroid of the target domain is calculated using the updated instance feature in real time.
- [00179] The self-spaced contrastive learning strategy described in the embodiments of the disclosure follows the principle of "from simple to difficult". The most trusted cluster

is learned, and then more trusted clusters may be gradually obtained to improve the quality of the learning target, so that errors can be reduced by increasing the number of the trusted clusters. The strategy provides a clustering credibility evaluation criterion. Evaluation of the independence and density of the cluster may be performed to select and retain the most trusted cluster and classify other clusters as non-cluster samples to provide instance-level supervision.

[00180] The unified contrastive learning framework is trained mainly through the following two steps that are under alternate execution.

[00181] Unlabeled target-domain samples are divided into two parts, i.e., a cluster set and a non-cluster set, by clustering and a clustering credibility evaluation criterion, to provide a cluster-level supervision and a non-cluster instance-level supervision respectively.

[00182] Training is performed through a proposed unified contrastive learning loss based on the source-domain class-level, target-domain cluster-level and target-domain non-cluster instance-level supervisions provided by the hybrid memory module, thereby optimizing the picture coder. The picture feature generated by the picture coder is used to dynamically update the hybrid memory module, a source-domain picture is updated by taking a class as a unit, and a target-domain picture is updated by taking an instance as a unit.

[00183] The embodiments of the disclosure provide a unified contrastive learning framework, and all training samples of a source domain and a target domain may be simultaneously learned to achieve advanced performance. The embodiments of the disclosure also provide a self-spaced learning strategy, and the clustering credibility evaluation criterion is provided to reduce training errors by trusted clustering. In a DA learning process, the performance of a source domain may be also improved. Class-level, cluster-level and instance-level supervisions may be simultaneously provided through a unified contrastive learning loss function. For unsupervised DA for pedestrian re-ID and vehicle re-ID, the most advanced ID effect is achieved. Training may be performed more effectively with unlabeled target-domain data to improve the performance of the labeled source domain. The training set may be extended using unlabeled data to improve the training performance.

[00184] Feature information of a target picture may be extracted using the picture coder of the algorithm in the embodiments of the disclosure. A pedestrian or vehicle in a security monitoring scenario may be retrieved using the feature extracted by the algorithm of the embodiments of the disclosure. The capability of the picture coder may be improved under an unsupervised condition using the algorithm of the embodiments of the disclosure.

[00185] FIG. 9 is a schematic diagram of a method for training a re-ID network by semi-supervised learning according to an embodiment of the disclosure. Referring to FIG. 9, the method for training the re-ID network includes the following steps.

[00186] In S901, a ResNet (initial network) 901 is acquired.

[00187] In S902, a first picture dataset 9021 and a second picture dataset 9022 are acquired from a hybrid memory 902, the first picture dataset 9021 including unlabeled target-domain picture data X^t , and the second picture dataset 9022 including source-domain picture data X^s including a true cluster label. The source-domain picture data is also called second cluster picture data.

[00188] In S903, clustering is performed on the target-domain picture data X^t in the first picture dataset through the ResNet 901 to obtain an initial clustering result, the initial clustering result including initial cluster picture data and initial non-cluster picture data.

[00189] In S904, re-clustering is performed on the initial clustering result to obtain first cluster picture data and non-cluster instance picture data.

[00190] In S905, a picture data center is determined based on training picture data.

[00191] The training picture data includes the first cluster picture data, the non-cluster instance picture data and the second cluster picture data. The picture data center includes a first cluster center corresponding to the first cluster picture data, an instance center corresponding to the non-cluster instance picture data and a second cluster center corresponding to the second cluster picture data. All the first cluster center, second cluster center and instance center that may be determined may be stored in the hybrid memory 902.

- [00192] In some embodiments, S905 may include the following steps.
- [00193] In S9051, a contrastive loss is determined based on the training picture data and the picture data center, and parameter optimization is performed on the ResNet 901 based on the contrastive loss to obtain an optimal network.
- [00194] In S9052, the non-cluster instance picture data in the training picture data is clustered through the optimal network, and the first cluster picture data and non-cluster instance picture data in the hybrid memory 902 are updated based on a clustering result to obtain new training picture data f^s and f^t , f^s including the second cluster picture data, and f^t including updated first cluster picture data and non-cluster instance picture data.
- [00195] In S9053, a new picture data center is determined based on the new training picture data, and a step that a new contrastive loss is determined based on the new training picture data and the new picture data center is repeated until training is completed, to obtain the re-ID network.
- [00196] The hybrid memory 902 may be updated according to the new training data f^s and f^t .
- [00197] In some embodiments, the operation in S904 that re-clustering is performed on the initial clustering result, referring to FIG. 10a, may include the following steps.
- [00198] In S9041, a picture data volume of a first present cluster in the initial cluster picture data is reduced based on a picture feature distance to obtain a second present cluster.
- [00199] Referring to FIG. 10a, the dots may represent picture data, the white dots may represent the initial cluster picture data, and the gray dots may represent the initial cluster picture data. If the picture feature distance changes from d_1 to d_2 , and $d_2 < d_1$, a picture feature distance between picture data 1011a and picture data 1012a in the first present cluster 101a is greater than d_2 , so that the two picture data is removed from the first present cluster 101a to reduce picture data in the first present cluster 101a to obtain the new second present cluster 102a.

- [00200] In S9042, a density index of the second present cluster is determined, the density index being a ratio of the picture data volume of the second present cluster to the picture data volume of the first present cluster.
- [00201] Referring to FIG. 10a, the picture data volume of the second present cluster is 5, and the picture data volume of the first present cluster is 7. In such a case, the density index of the second present cluster is $5/7$.
- [00202] In S9043, when the density index reaches a first preset threshold, the first present cluster is replaced with the second present cluster to obtain the first cluster picture data 90211.
- [00203] If the first preset threshold is 0.5, the density index is greater than the first preset threshold, so that the first cluster picture data may be the picture data in the second present cluster 102a.
- [00204] In S9044, clipped picture data is classified as non-cluster instance picture data 90212.
- [00205] Referring to FIG. 10a, the clipped picture data 1011a and picture data 1012a may be classified as non-cluster instance picture data 90212. In such a case, the non-cluster instance picture data includes the initial non-cluster picture data represented by the gray dots, as well as the picture data 1011a and the picture data 1012a.
- [00206] In some embodiments, the operation in S904 that re-clustering is performed on the initial clustering result, referring to FIG. 10b, may include the following steps.
- [00207] In S9045, picture data of another cluster and/or picture data in the initial non-cluster picture data are/is added to a third present cluster of the initial cluster picture data based on the picture feature distance to obtain a fourth present cluster, the another cluster being different from the third present cluster in the initial cluster picture data.
- [00208] Referring to FIG. 10b, the dots may represent picture data, the white dots may represent the initial cluster picture data, and the gray dots may represent the initial non-cluster picture data. For the third present cluster 101b and other cluster 102b existing before re-clustering, if the picture feature distance changes from $d1$ to $d3$, and $d3 > d1$, picture feature distances of initial non-cluster picture data 1011b, initial non-cluster

picture data 1012b and initial non-cluster picture data 1013b are less than d3, so that the initial non-cluster picture data 1011b, the initial non-cluster picture data 1012b, the initial non-cluster picture data 1013b and the picture data in the another cluster 102b are added to the third present cluster 101b to increase the picture data volume of the third present cluster 101b to obtain the new fourth present cluster 1013b.

[00209] In S9046, an independence index of the fourth present cluster is determined, the independence index being a ratio of a picture data volume of the third present cluster to a picture data volume of the fourth present cluster.

[00210] Referring to FIG. 10b, the picture data volume of the third present cluster 101b is 3, and the picture data volume of the fourth present cluster 103b is 9. In such a case, the independence index of the fourth present cluster 103b is 3/9.

[00211] In S9047, when the independence index reaches the first preset threshold, the third present cluster is replaced with the fourth present cluster to obtain the first cluster picture data.

[00212] When the first preset threshold is 0.3, the independence index is greater than the first preset threshold, so that the first cluster picture data 90211 may be the picture data in the fourth present cluster 103b.

[00213] In S9048, when added picture data includes the picture data of the another cluster, the another cluster is canceled, and/or, when the added picture data includes the picture data in the initial non-cluster picture data, the added picture data is not classified as non-cluster instance picture data 90212.

[00214] The another cluster 102b and the third cluster 101b may be canceled. Since the added picture data includes the picture data 1011b, picture data 1012b and picture data 1013b in the initial cluster picture data, the picture data 1011b, the picture data 1012b and the picture data 1013b are not classified as non-cluster instance picture data, namely the non-cluster instance picture data excludes the picture data 1011b, the picture data 1012b and the picture data 1013b.

[00215] It should be understood that, although each step in the flowcharts of the embodiments is sequentially presented according to the indications of the arrowheads,

these steps are not always executed according to sequences indicated by the arrowheads. Unless otherwise clearly described in the disclosure, no strict limits are made to execution sequences of these steps, and these steps may be executed in other sequences. Moreover, at least part of steps in the figures may include multiple sub-steps or multiple stages, these sub-steps or stages are not always executed and completed at the same time but may be executed at different time, and they are not always sequentially executed but may be executed in turn or alternately with at least part of other steps or sub-steps or stages of the other steps.

[00216] In an embodiment, there is provided a device for training a re-ID network.

[00217] FIG. 11 is a schematic diagram of a device for training a re-ID network according to an embodiment of the disclosure. As shown in FIG. 11, the device includes a first acquisition module 100, a second acquisition module 200 and a network training module 300.

[00218] The first acquisition module 100 is configured to acquire an initial network.

[00219] The second acquisition module 200 is configured to acquire training picture data.

[00220] The network training module 300 is configured to train the initial network through the training picture data to obtain a re-ID network.

[00221] Limits to the device for training a re-ID network may refer to the above limits to the method for training a re-ID network, and are not elaborated here. The modules in the device for training a re-ID network may completely or partially be implemented by software, hardware, and a combination thereof. Each module may be embedded in a hardware form into or independent of a processor in a computer device, or may be stored in a software form in memory in the computer device, for the processor to call to execute the operations corresponding to each module.

[00222] In an embodiment, there is provided a device for re-identifying an object.

[00223] FIG. 12 is a schematic diagram of a device for re-identifying an object according to an embodiment of the disclosure. As shown in FIG. 12, the device includes a network acquisition module 400, a picture acquisition module 500 and a re-ID module

600.

[00224] The network acquisition module 400 is configured to acquire a pre-trained re-ID network.

[00225] The picture acquisition module 500 is configured to acquire a picture to be identified.

[00226] The re-ID module 600 is configured to perform re-ID on the picture to be identified through the re-ID network to obtain a re-ID result of a target object in the picture to be identified.

[00227] Training picture data for the re-ID network at least includes first cluster picture data and non-cluster instance picture data. The first cluster picture data and the non-cluster instance picture data are obtained by performing clustering on a first picture dataset through an initial network corresponding to the re-ID network, and picture data in the first picture dataset exclude true cluster labels.

[00228] In an embodiment, the training picture data for the re-ID network may further include a second picture dataset. Second cluster picture data in the second picture dataset include a true cluster label, and a picture data domain of the second picture dataset is different from a picture data domain of the first picture dataset.

[00229] In an embodiment, the device may further include: an initial network acquisition module, configured to acquire the initial network; a data acquisition module, configured to acquire the training picture data; and a training module, configured to train the initial network through the training picture data to obtain the re-ID network.

[00230] In an embodiment, the data acquisition module may include: a result acquisition unit, configured to acquire an initial clustering result obtained by performing clustering on the first picture dataset through the initial network; and a clustering unit, configured to perform re-clustering on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data.

[00231] In an embodiment, the initial clustering result may include initial cluster picture data, and the clustering unit is configured to: reduce a picture data volume of a first present cluster in the initial cluster picture data based on a picture feature distance to

obtain a second present cluster; determine a density index of the second present cluster, the density index being a ratio of a picture data volume of the second present cluster to the picture data volume of the first present cluster; when the density index reaches a first preset threshold, replace the first present cluster with the second present cluster to obtain the first cluster picture data; and classify clipped picture data as the non-cluster instance picture data.

[00232] In an embodiment, the initial clustering result may further include initial non-cluster picture data. The clustering unit is further configured to: add picture data of another cluster and/or picture data in the initial non-cluster picture data to a third present cluster of the initial cluster picture data based on the picture feature distance to obtain a fourth present cluster, the another cluster being different from the third present cluster in the initial cluster picture data; determine an independence index of the fourth present cluster, the independence index being a ratio of a picture data volume of the third present cluster to a picture data volume of the fourth present cluster; when the independence index reaches the first preset threshold, replace the third present cluster with the fourth present cluster to obtain the first cluster picture data; and when the added picture data includes the picture data of the another cluster, cancel the another cluster, and/or, when the added picture data includes the picture data in the initial non-cluster picture data, not classify the added picture data as non-cluster instance picture data.

[00233] In an embodiment, the training module may include: a first determination unit, configured to determine a picture data center based on the training picture data; an optimization unit, configured to determine a contrastive loss based on the training picture data and the picture data center and perform parameter optimization on the initial network based on the contrastive loss to obtain an optimal network; a clustering unit, configured to cluster the non-cluster instance picture data in the training picture data through the optimal network and update the first cluster picture data and the non-cluster instance picture data based on a clustering result to obtain new training picture data; and a second determination unit, configured to determine a new picture data center based on the new training picture data and repeat a step of determining a new contrastive loss based on the new training picture data and the new picture data center until training is completed, to obtain the re-ID network.

[00234] In an embodiment, the picture data center may include a first cluster center

corresponding to the first cluster picture data and an instance center corresponding to the non-cluster instance picture data; or, the picture data center may include a first cluster center corresponding to the first cluster picture data, an instance center corresponding to the non-cluster instance picture data and a second cluster center corresponding to the second cluster picture data.

[00235] In an embodiment, the re-ID network may include a ResNet.

[00236] Limits to the device for re-identifying an object may refer to the above limits to the method for re-identifying an object, and are not elaborated here. The modules in the device for re-identifying an object may completely or partially be implemented by software, hardware, and a combination thereof. Each module may be embedded in a hardware form into or independent of a processor in a computer device, or may be stored in a software form in memory in the computer device, for the processor to call to execute the operations corresponding to each module.

[00237] The embodiments of the disclosure also provide a computer program product including a non-transitory computer-readable storage medium storing a computer program which enables a computer to execute part or all of the steps of any method for re-identifying an object recorded in the method embodiments.

[00238] In an embodiment, there is provided a computer device, which includes memory, a processor and a computer program stored in the memory and capable of running in the processor. The processor executes the program to implement the method steps of the network training part and/or the method steps of the network application part in each of the above embodiments.

[00239] In an embodiment, there is provided a computer-readable storage medium having stored therein a computer-executable instruction which, when executed by a processor, is configured to implement the method steps of the network training part and/or the method steps of the network application part in each of the above embodiments.

[00240] Other implementation solutions of the disclosure are apparent to those skilled in the art after considering the specification and practicing the application disclosed here. The disclosure is intended to cover any transformations, uses or adaptive variations of the

disclosure, and these transformations, uses or adaptive variations follow the general principle of the disclosure, and include common general knowledge or conventional technical means undisclosed in the disclosure in the technical field. The specification and the embodiments are only considered as examples, and the practical scope and spirit of the disclosure are pointed out by the appended claims.

[00241] It should be understood that the disclosure is not limited to the precise structures described above and shown in the drawings, and various modifications and variations may be made without departing from the scope thereof. The scope of the disclosure is only defined by the appended claims.

INDUSTRIAL APPLICABILITY

[00242] In the embodiments, a pre-trained re-ID network is acquired; a picture to be identified is acquired; and re-ID is performed on the picture to be identified through the re-ID network to obtain a re-ID result of a target object in the picture to be identified. The re-ID network used in the method may be obtained by training at least first cluster picture data and non-cluster instance picture data. Therefore, in the disclosure, network training is performed in combination with outliers not included in clusters to improve the clustering performance of the re-ID network and further improve the accuracy of the re-ID result of the target object obtained through the re-ID method of the disclosure.

CLAIMS

1. A method for re-identifying an object, comprising:

acquiring a pre-trained re-Identification (ID) network;

acquiring a picture to be identified; and

performing re-ID on the picture to be identified through the re-ID network to obtain a re-ID result of a target object in the picture to be identified,

wherein training picture data for the re-ID network at least comprise first cluster picture data and non-cluster instance picture data, the first cluster picture data and the non-cluster instance picture data are obtained by performing clustering on a first picture dataset through an initial network corresponding to the re-ID network, and picture data in the first picture dataset exclude true cluster labels.

2. The method of claim 1, wherein the training picture data for the re-ID network further comprise a second picture dataset, and second cluster picture data in the second picture dataset comprise a true cluster label; and

a picture data domain of the second picture dataset is different from a picture data domain of the first picture dataset.

3. The method of claim 1 or 2, before acquiring the pre-trained re-ID network, further comprising:

acquiring the initial network;

acquiring the training picture data; and

training the initial network through the training picture data to obtain the re-ID network.

4. The method of claim 3, wherein acquiring the training picture data comprises:

acquiring an initial clustering result obtained by performing clustering on the first picture dataset through the initial network; and

performing re-clustering on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data.

5. The method of claim 4, wherein the initial clustering result comprises initial cluster picture data,

performing re-clustering on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data comprises:

reducing a picture data volume of a first present cluster in the initial cluster picture data based on a picture feature distance to obtain a second present cluster;

determining a density index of the second present cluster, the density index being a ratio of a picture data volume of the second present cluster to the picture data volume of the first present cluster;

when the density index reaches a first preset threshold, replacing the first present cluster with the second present cluster to obtain the first cluster picture data; and

classifying clipped picture data as non-cluster instance picture data.

6. The method of claim 5, wherein the initial clustering result further comprises initial non-cluster picture data,

performing re-clustering on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data comprises:

adding picture data of another cluster and/or picture data in the initial non-cluster

picture data to a third present cluster of the initial cluster picture data based on the picture feature distance to obtain a fourth present cluster, the another cluster being different from the third present cluster in the initial cluster picture data;

determining an independence index of the fourth present cluster, the independence index being a ratio of a picture data volume of the third present cluster to a picture data volume of the fourth present cluster;

when the independence index reaches the first preset threshold, replacing the third present cluster with the fourth present cluster to obtain the first cluster picture data; and

when the added picture data includes the picture data of the another cluster, canceling the another cluster, and/or, when the added picture data includes the picture data in the initial non-cluster picture data, not classifying the added picture data as non-cluster instance picture data.

7. The method of claim 3, wherein training the initial network through the training picture data to obtain the re-ID network comprises:

determining a picture data center based on the training picture data;

determining a contrastive loss based on the training picture data and the picture data center, and performing parameter optimization on the initial network based on the contrastive loss to obtain an optimal network;

clustering the non-cluster instance picture data in the training picture data through the optimal network, and updating the first cluster picture data and the non-cluster instance picture data based on a clustering result to obtain new training picture data; and

determining a new picture data center based on the new training picture data, and repeating a step of determining a new contrastive loss based on the new training picture data and the new picture data center until training is completed, to obtain the re-ID network.

8. The method of claim 7, wherein

the picture data center comprises a first cluster center corresponding to the first cluster picture data and an instance center corresponding to the non-cluster instance picture data; or,

the picture data center comprises a first cluster center corresponding to the first cluster picture data, an instance center corresponding to the non-cluster instance picture data and a second cluster center corresponding to the second cluster picture data.

9. The method of any one of claims 1-8, wherein the re-ID network comprises a Residual Network (ResNet).

10. A device for re-identifying an object, comprising:

a network acquisition module, configured to acquire a pre-trained re-Identification (ID) network;

a picture acquisition module, configured to acquire a picture to be identified; and

a re-ID module, configured to perform re-ID on the picture to be identified through the re-ID network to obtain a re-ID result of a target object in the picture to be identified,

wherein training picture data for the re-ID network at least comprise first cluster picture data and non-cluster instance picture data, the first cluster picture data and the non-cluster instance picture data are obtained by performing clustering on a first picture dataset through an initial network corresponding to the re-ID network, and picture data in the first picture dataset exclude true cluster labels.

11. The device of claim 10, wherein the training picture data for the re-ID network further comprise a second picture dataset, and second cluster picture data in the second picture dataset comprise a true cluster label; and

a picture data domain of the second picture dataset is different from a picture data domain of the first picture dataset.

12. The device of claim 10 or 11, further comprising:

an initial network acquisition module, configured to acquire the initial network;

a data acquisition module, configured to acquire the training picture data; and

a training module, configured to train the initial network through the training picture data to obtain the re-ID network.

13. The device of claim 12, wherein the data acquisition module comprises:

a result acquisition unit, configured to acquire an initial clustering result obtained by performing clustering on the first picture dataset through the initial network; and

a clustering unit, configured to perform re-clustering on the initial clustering result to obtain the first cluster picture data and the non-cluster instance picture data.

14. The device of claim 13, wherein the initial clustering result comprises initial cluster picture data,

the clustering unit is configured to:

reduce a picture data volume of a first present cluster in the initial cluster picture data based on a picture feature distance to obtain a second present cluster;

determine a density index of the second present cluster, the density index being a ratio of a picture data volume of the second present cluster to the picture data volume of the first present cluster;

when the density index reaches a first preset threshold, replace the first present cluster with the second present cluster to obtain the first cluster picture data; and

classify clipped picture data as non-cluster instance picture data.

15. The device of claim 14, wherein the initial clustering result further comprises initial non-cluster picture data,

the clustering unit is further configured to:

add picture data of another cluster and/or picture data in the initial non-cluster picture data to a third present cluster of the initial cluster picture data based on the picture feature distance to obtain a fourth present cluster, the another cluster being different from the third present cluster in the initial cluster picture data;

determine an independence index of the fourth present cluster, the independence index being a ratio of a picture data volume of the third present cluster to a picture data volume of the fourth present cluster;

when the independence index reaches the first preset threshold, replace the third present cluster with the fourth present cluster to obtain the first cluster picture data; and

when the added picture data includes the picture data of the another cluster, cancel the another cluster, and/or, when the added picture data includes the picture data in the initial non-cluster picture data, not classify the added picture data as non-cluster instance picture data.

16. The device of claim 12, wherein the training module comprises:

a first determination unit, configured to determine a picture data center based on the training picture data;

an optimization unit, configured to determine a contrastive loss based on the training picture data and the picture data center and perform parameter optimization on the initial network based on the contrastive loss to obtain an optimal network;

a clustering unit, configured to cluster the non-cluster instance picture data in the training picture data through the optimal network and update the first cluster picture data and the non-cluster instance picture data based on a clustering result to obtain new training picture

data; and

a second determination unit, configured to determine a new picture data center based on the new training picture data and repeat a step of determining a new contrastive loss based on the new training picture data and the new picture data center until training is completed, to obtain the re-ID network.

17. The device of claim 16, wherein

the picture data center comprises a first cluster center corresponding to the first cluster picture data and an instance center corresponding to the non-cluster instance picture data; or,

the picture data center comprises a first cluster center corresponding to the first cluster picture data, an instance center corresponding to the non-cluster instance picture data and a second cluster center corresponding to the second cluster picture data.

18. The device of any one of claims 10-17, wherein the re-ID network comprises a Residual Network (ResNet).

19. A computer device, comprising memory, a processor and a computer program stored in the memory and capable of running in the processor, the processor executing the program to implement the steps of the method for re-identifying an object according to any one of claims 1-9.

20. A computer-readable storage medium having stored therein a computer-executable instruction which, when executed by a processor, implements the method for re-identifying an object according to any one of claims 1-9.

21. A computer program product comprising a computer-readable code that, when runs in an

electronic device, causes a processor in the electronic device to execute the method for re-identifying an object according to any one of claims 1-9.