(10) Publication No: **IN585888 B1**

(45) Date of Grant: 30-03-2026

(19) Intellectual Property Office, India

(21) Application No: 202141034299

(24) Date of Patent: 30-07-2021

(12) INDIAN GRANTED PATENT

(54) Title: **A system and method for image content translation across different cameras**

(51) International Classification:
G06N3/08, G06K9/62, G06N3/04, G06T7/593, G06F16/182

(22) Date of Filing: **30-07-2021**

(43) Publication Date: **03-02-2023**
Journal No: **05/2023**

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

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

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

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

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

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

(71) Name of Applicant(s):
1. Bosch Global Software Technologies Private Limited,
2. Robert Bosch Gmbh,

(72) Name of Inventor(s):
1. Amit Arvind Kale
2. Harshil Jain
3. Koustav Mullick

(57) Abstract:

A SYSTEM AND METHOD FOR IMAGE CONTENT TRANSLATION ACROSS DIFFERENT CAMERAS ABSTRACT Method (200) and system (100) for domain adaptation for semantic similar images using generative adversarial network (GAN) are described. A plurality of images (102) acquired from two or more different imaging devices are received as an input to an image processing module (104) for identifying a semantically similar pair of images (106) from the received plurality of images (102). The identified semantically similar pair of images (106) are from a source domain and a target domain. The identified semantically similar pair of images (106) from said image processing module (104) are obtained by a generative adversarial network (GAN) (116). The generative adversarial network (GAN) (116) is trained to perform image translation between source domain images and target domain images to generate a domain translated output. Further, domain translated output from the GAN (116) is compared with images from both the source domain and the paired target domain by using a comparing module (108), in order to ensure better consistency across the domains. (Figure 1)

FORM 2

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

COMPLETE SPECIFICATION
(SECTION 10 and Rule 13)

1. **TITLE OF THE INVENTION: A SYSTEM AND METHOD FOR IMAGE CONTENT TRANSLATION ACROSS DIFFERENT CAMERAS**

2. **APPLICANTS:**

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

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

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

FIELD OF THE INVENTION

[0001] The present subject matter relates, in general, to a system and method for improving domain adaptation image translation task in a network to process images differ in visual properties and acquired from different imaging devices(e.g. cameras), using Generative Adversarial Networks (GANs).

BACKGROUND OF THE INVENTION

[0002] A Generative Adversarial Network, or GAN, is a type of neural network architecture for generative modeling. Generative modeling involves using a model to generate new examples that plausibly come from an existing distribution of samples, such as generating new photographs that are similar but specifically different from a dataset of existing photographs. A GAN is a generative model that is trained using two neural network models. One model is called the “generator” or “generative network” model that learns to generate new plausible samples. The other model is called the “discriminator” or “discriminative network” and learns to differentiate generated examples from real examples.

[0003] While humans can easily translate an image into another image, for instance, imagining the missing region of corrupted image or sketching the scenic photograph, it is challenging for machine to automatically learn the mapping especially when supervision is absent. Similar tasks include image colorization, image semantic segmentation, and image denoising. All of these can be framed as image-to-image translation as they could be formulated as pixel regression or classification. However, these methods are all pixel-wise matching oriented, being inadequate for high-level representations and tend to yield blurry outputs.

[0004] Recent work moved beyond specific image translation tasks by developing a GANs-based common framework for various image translation tasks. Such method requires explicitly aligned data in the form of input image for training, which is seldom available in practice. Attempts to alleviate this issue had also been made to exploit dual learning with cyclic loss. However, cycle-consistency enforced via or loss tends to induce the averaging of potential location of details and thus leads to over smoothed images.

[0005] Domain adaptation is proposed to learn representations that are invariant to data from different distributions. In other words, cross-domain relation in the form of a mapping from source domain to target domain needs to be built, which is quite similar to the goal in image-to-image translation. For autonomous vehicles, perception tasks contribute highly for ability of autonomous vehicle to collect information and knowledge about the external world. It helps in developing a contextual understanding about important signs like where drivable space is, where obstacles are, where traffic signs and signals are and what they are specifying about the scene. All of these is required to be understood and then decide upon the trajectory of the vehicle and how control of the vehicle will happen over next instances.

[0006] In this case, we are assuming the perception block to be made of cameras. Cameras are used for sensing this information. With these cameras, data is captured and is annotated with various classes for performing complex computer vision tasks like segmentation, detection etc. One of the widely used approaches to learn these tasks is with the help of deep-learning models which rely heavily on data and more the data, the higher the power of models to be able to generalize.

[0007] But over the time, with improvements in technology, the cameras used in sensing systems also get upgraded like higher resolution, better sensing ability, improved night-time vision, higher field-of-view and likewise. But it also brings about a shift in the image properties because of the shift in distribution between the previous and newer sensors of the cameras.

[0008] In order to circumvent the above mentioned possible drawback and prevent the discriminator from picking up on niche queues of semantic differences, it is important to enforce consistency across the image pairs occurring in a particular batch. Semantic inconsistencies across batches of the two domains while training an image translation network could potentially lead to unstable training of the discriminator.

[0009] A prior art, WO2017158363A1 discloses a method for training an algorithm to process at least a section of received visual data using a training dataset and reference dataset. The method comprises an iterative method with each iteration comprising the steps of: generating a set of training data using the algorithm; comparing one or more characteristics of the training data to one or more characteristics of at least a section of the reference dataset; and modifying one or more parameters of the algorithm to optimize processed visual data based on the comparison between the characteristic of the training data and the characteristic of the reference dataset. The algorithm may output the processed visual data with the same content as the at least a section of received visual data. Some aspects and/or embodiments provide for improved super-resolution of lower quality images, with a view to producing super-resolution images which have improved characteristics (e.g. less blur, less undesired smoothing) compared to other super-resolution techniques.

BRIEF DESCRIPTION OF THE ACCOMPANYING DRAWING

[0010] The detailed description is provided with reference to the accompanying figures, wherein:

[0011] FIG. 1 illustrates a system environment for domain adaptation for semantic similar images using generative adversarial network (GAN), in accordance with an example implementation of the present subject matter;

[0012] FIG. 2 illustrates a flow chart of a method for domain adaptation for semantic similar images using generative adversarial network (GAN), in accordance with an example implementation of the present subject matter; and

[0013] FIG's. 3A & 3B illustrate, paired set of images obtained using the proposed system environment for domain adaptation from two reference imaging devices (e.g. cameras) placed on vehicle, in accordance with an example implementation of the present subject matter.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0014] FIG. 1 illustrates a system environment for domain adaptation image translation task for images acquired from different imaging devices using generative adversarial network (GAN), in accordance with an example implementation of the present subject matter. Herein, the different imaging devices may include cameras of different generations installed on a vehicle. The proposed system environment helps to extract out pair of images corresponding to two available image streams from different cameras.

[0015] The system environment may include a computing system 100 and a neural network architecture. The computing system 100 may be communicatively coupled to the generative adversarial network (GAN) architecture. In an example, the computing system 100 may be directly or remotely coupled to the generative adversarial network (GAN) architecture. Examples of the computing system 100 may include, but are not limited to, a laptop, a notebook computer, a desktop computer, and so on.

[0016] The computing system 100 may include a memory 110. The memory example, volatile memory, such as static random-access memory (SRAM) and dynamic random-access memory (DRAM), and/or non-volatile memory, such as read only memory (ROM), erasable programmable ROM, flash memories, hard disks, optical disks, and magnetic tapes.

[0017] In an example, the computing system 100 may also include a processor 112 coupled to the memory 110. The processor 112 may include microprocessors, microcomputers, microcontrollers, digital signal processors, central processing units, state machines, logic circuitries, and/or any other devices that manipulate signals and data based on computer-readable instructions. Further, functions of the various elements shown in the figures, including any functional blocks labelled as “processor(s)”, may be provided through the use of dedicated hardware as well as hardware capable of executing computer-readable instructions.

[0018] Further, the computing system 100 may include interface(s) 114. The interface(s) 114 may include a variety of interfaces, for example, interface(s) for users. The interface(s) 114 may include data output devices. In an example, the interface(s) 114 may provide an interactive platform for receiving inputs from a user. For example, the user may provide a plurality of images 102 from acquired from different imaging devices as an input to the computing system 100 through the interface(s) 114. In an example, the plurality of images acquired from different imaging devices may be images of driving scenes dataset which consists of images containing different set of objects like vehicles, person, road, buildings and sky, in the wild.

[0019] Furthermore, the processor 112 of the computing system 100 is configured to receive a plurality of images 102 acquired from two or more different imaging devices and transmit the plurality of images 102 to an image processing module 104. The image processing module 104 is configured to process the plurality of images 102, to identify a semantically similar pair of images 106. Herein, the pair of images 106 are from a source domain and a target domain. In one exemplary embodiment, the pair of images 102 from two or more domains may include images acquired by using different imaging sensors, which differs in visual properties such as pixel intensity range, field of view, resolution and the like.

[0020] The image processing module 104 is communicatively coupled to a generative adversarial network (GAN) 116. The image processing module 104 is further configured to transmit identified semantically similar pair of images 106 to the generative adversarial network (GAN) (116). The GAN 116 is trained to perform image translation between said source domain and target domain to generate a domain translated output image.

[0021] In general, the GAN framework is trained for image translation without using the paired data into input account. It works on the principle of Generative Adversarial networks which has two main components, namely generator and discriminator which competes for improving the performance of overall network. The role of the generator is to figure out translation of the image from source to target domain in such a manner that it is able to fool the discriminator, whose role is to act as critic and distinguish between original and translated images. However, in the proposed invention if the pair of images passed during the learning process of GAN, has semantic resemblance, the network learns to translate images with both improved perceptual quality and feature distribution adaptability. Availability of paired data acquired using the image processing module 104, enables the GAN 116 to learn the transformation system better. During learning, over time the discriminator can no longer differentiate between the translated and original images from each domain, meaning the generator has learnt the translation parameters well.

[0022] Further, a comparing module 108 is communicatively connected to the generative adversarial network (GAN) 116 over a communication network. The comparing module 108 is configured to receive the domain translated output image from the GAN 116. The comparing module 104 is further configured to compare domain translated output image from the GAN 116 with images from both the source domain and the paired target domain, in order to ensure better consistency across the domains. In one exemplary embodiment, the comparing module 108 is further configured to ensure feature distribution consistency across the translated output images across the source and target domains.

[0023] The feature may represent a set of characteristics, such as a shape, color profiles, texture pattern, or a combination thereof, associated with the received set of images. In an example, the image processing module 104 may generate feature vectors from the set of features. The feature vectors may be a representation of the set images.

[0024] In one embodiment, the comparing module 108 is configured to divide the domain translated output image, from GAN 116 and the source domain images into plurality of sub-images. Each of the plurality of sub-images are further processed by the comparing module 108, in order to extract feature from each of the sub-images. In this embodiment, the comparing module 108 is further configured to minimize the squared difference between the extracted features from the plurality of the sub-images of the domain adapted image and the source image, respectively.

[0025] While minimizing the content consistency loss between source image and its corresponding domain translated image computation at quadrant level is performed, instead of at the whole image level to better capture the semantic information present in the images. By doing so, the block information is contextualized and comparing it across the image. In one embodiment, the original content consistency loss is computed over the whole image using features extracted from a pre-trained ImageNet model. Whereas operating at a quadrant level provides better supervision on reduced semantic area which enables better loss propagation. This improves the overall transferability obtained using the comparing module 108 based domain adaptation process.

[0026] In a similar manner, the comparing module 108 is configured to divide the domain translated output image, from GAN (116) and the paired target domain images into plurality of sub-images. Each of the plurality of sub-images are further processed by the comparing module 108 in order to extract feature from each of the sub-images. In this embodiment, the comparing module 108 is further configured to minimize the squared difference between the extracted features from the plurality of the sub-images of the domain adapted image and the paired target domain image, respectively.

[0027] Additionally along with the paired target domain image, the comparing module 108 incorporate quadrant based cost functions which takes into account the lack of exact pixel to pixel correspondence between the pair of semantically similar images 106. This circumvents the error which would have been induced otherwise while calculating exact pixel level differences between the loosely paired images.

[0028] FIG. 2 illustrates a flow chart of a method 200 for domain adaptation for paired semantic similar images using generative adversarial network (GAN) 116 in accordance with an example implementation of the present subject matter. The method 200 may be implemented by the computing system 100 including the memory 110, the processor 112, and the interface(s) 114, of FIG. 1. Further, the computing system 100 may be communicatively coupled with the neural network architecture as described in FIG. 1. Although, the method 200 is described in context of the system that is similar to the computing system 100 of FIG. 1, other suitable devices or systems may be used for execution of the method 200.

[0029] At block 201, the method 200 may include receiving a plurality of images 102 acquired from two or more different imaging devices and transmit the plurality of images 102 to an image processing module 104. In one exemplary embodiment of the invention, the plurality of images 102 from two or more domains may include images acquired by using different imaging sensors, which differs in visual properties such as pixel intensity range, field of view, resolution and the like.

[0030] At block 202, the method 200 may include identifying a semantically similar pair of images 106 from the received plurality of images 102 by using an image processing module 104. Herein, the identified semantically similar pair of images 106 are from a source domain and a target domain.

[0031] In one embodiment, the image processing module 104 synchronizes the plurality of images 102 acquired from two or more different imaging devices. Herein, the different imaging devices may include cameras of different generations installed on a vehicle. In one exemplary embodiment, the imaging devices are the two independent camera sensors and establish the correspondence between the respective images. The image processing module 104 is configured to get past the difference in the capturing properties of the two imaging devices such as frames per second, resolution, image format, FOV, and synchronize the images. It makes use of the spatial and object properties present in the content of the respective stream of images to derive the correspondence between them. The identified paired set of images obtained using the proposed system environment for domain adaptation from two reference imaging devices (e.g. cameras) placed on vehicle can be shown in FIG.'s 3A and 3B.

[0032] At block 203, the method 200 may include obtaining the identified semantically similar pair of images 106 from said image processing module 104, by a generative adversarial network (GAN) 116. A block 204, the method 200 may include performing image translation between said first source and target domain to generate a domain translated output by using the generative adversarial network (GAN) 116.

[0033] Further, at block 205, the method 200 may include receiving and comparing the domain translated output from the GAN 116 with images from both the source domain and the paired target domain, in order to ensure better consistency across the domains, by using a comparing module 108.

[0034] In this block 205, the comparing module 108 is configured to divide the domain translated output image, from GAN 116 and the source domain images/ paired target domain image into plurality of sub-images. Each of the plurality of sub-images are further processed by the comparing module 108, in order to extract feature from each of the sub-images. In this embodiment, the comparing module 108 is further configured to minimize the squared difference between the extracted features from the plurality of the sub-images of the domain adapted image and the source image/ paired target domain image, respectively.

[0035] While minimizing the content consistency loss between source image and its corresponding domain translated image computation at quadrant level is performed by the comparing module 108, instead of at the whole image level to better capture the semantic information present in the images. By doing so, the block information is contextualized and comparing it across the image. In one embodiment, the original content consistency loss is computed over the whole image using features extracted from a pre-trained ImageNet model. Whereas operating at a quadrant level provides better supervision on reduced semantic area which enables better loss propagation. This improves the overall transferability obtained using the comparing module 108 based domain adaptation process.

[0036] In one embodiment, translated images resemble each other in the translated domain can be achieved by a cost function in the GAN 116 framework to ensure that the feature distance between the set of images from multiple domain can be minimized. In this example, the feature distance, the Euclidean distance can be considered between the respective pairs of semantic similar images 106. Additionally two new cost functions can be added in the GAN 116 to minimize the transformed images to further exploit the pairing of semantic similar images 106. In one embodiment, the method 200 continues to iterate until the GAN 116 produces visually better quality translated images from synthetic to real-world domain

[0037] The proposed present invention gives an approach to capture and thereby reduce the distribution shift between the domains. This approach takes only the images of two domains and does not require any annotation into account for learning the translation of images from one domain to another.

[0038] Although aspects for the present disclosure have been described in a language specific to structural features and/or methods, it is to be understood that the appended claims are not limited to the specific features or methods described herein. Rather, the specific features and methods are disclosed as examples of the present disclosure.

CLAIMS

We Claim:

1. A computing system (100) comprising:
   a memory (110); and
   a processor (112), coupled to the memory (110), to receive a plurality of images (102) acquired from two or more different imaging devices and transmit the plurality of images (102) to an image processing module (104);
   said image processing module (104) is configured to process the plurality of images (102) to identify a semantically similar pair of images (106), wherein said semantically similar pair of images (106) are from a source domain and a target domain,
   wherein said image processing module (104) is communicatively coupled to a generative adversarial network (GAN) (116) and configured to transmit identified semantically similar pair of images (106) to the generative adversarial network (GAN) (116) , which is trained to perform image translation between said source domain and target domain to generate a domain translated output image;
   characterized in that, a comparing module (108) is configured to receive and compare the domain translated output image from the GAN (116) with images from both the source domain and the paired target domain, in order to ensure better consistency across the domains.

2. The computing system (100) as claimed in claim 1, wherein said comparing module (108) is further configured to:
   divide the domain translated output image, from GAN (116) and the source domain images into plurality of sub-images;

process each of the plurality of sub-images in order to extract feature from each of the sub-images; and

minimize the squared difference between the extracted features from the sub-images of the domain adapted image and the source image, respectively.

3. The computing system (100) as claimed in claim 1, wherein said comparing module (108) is further configured to:

divide the domain translated output image, from GAN (116) and the paired target domain images into plurality of sub-images;

process each of the plurality of sub-images in order to extract feature from each of the sub-images; and

minimize the squared difference between the extracted features from the sub-images of the domain adapted image and the paired target domain images, respectively.

4. The computing system (100) as claimed in claim 1, wherein said comparing module (108) is further configured to ensure feature distribution consistency across the translated output images across the source and target domains.

5. The computing system (100) as claimed in claim 1, wherein the different imaging devices may include cameras of different generations installed on a vehicle.

6. A method (200) for domain adaptation for domain adaptation for semantic similar images using generative adversarial network (GAN), the method (200) comprising the steps for:

receiving (201) a plurality of images (102) acquired from two or more different imaging devices by an image processing module (104);

identifying (202) a semantically similar pair of images (106) from the received plurality of images (102) by using an image processing module (104), wherein said semantically similar pair of images (106) are from a source domain and a target domain;

obtaining (203) the identified semantically similar pair of images (106) from said image processing module (104) by a generative adversarial network (GAN) (116);

performing (204) image translation between said source domain and target domain to generate a domain translated output by using said generative adversarial network (GAN) (116);

receiving and comparing (205) the domain translated output from the GAN (116) with images from both the source domain and the paired target domain, in order to ensure better consistency across the domains, by using a comparing module (108).

Dated this 30th day of July 2021

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

A SYSTEM AND METHOD FOR IMAGE CONTENT TRANSLATION ACROSS DIFFERENT CAMERAS

ABSTRACT

Method (200) and system (100) for domain adaptation for semantic similar images using generative adversarial network (GAN) are described. A plurality of images (102) acquired from two or more different imaging devices are received as an input to an image processing module (104) for identifying a semantically similar pair of images (106) from the received plurality of images (102). The identified semantically similar pair of images (106) are from a source domain and a target domain. The identified semantically similar pair of images (106) from said image processing module (104) are obtained by a generative adversarial network (GAN) (116). The generative adversarial network (GAN) (116) is trained to perform image translation between source domain images and target domain images to generate a domain translated output. Further, domain translated output from the GAN (116) is compared with images from both the source domain and the paired target domain by using a comparing module (108), in order to ensure better consistency across the domains.

(Figure 1)

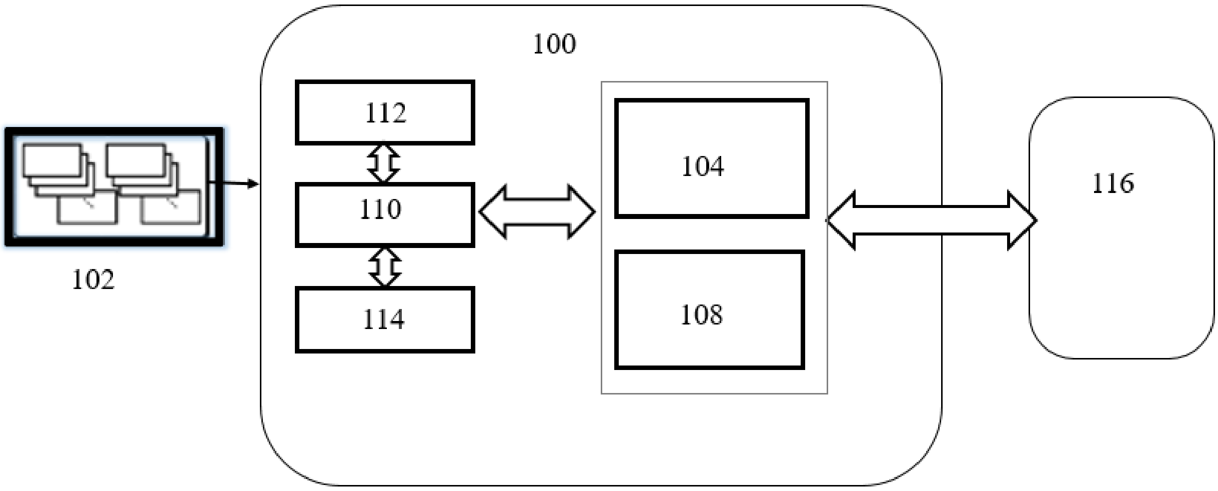

Figure 1

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

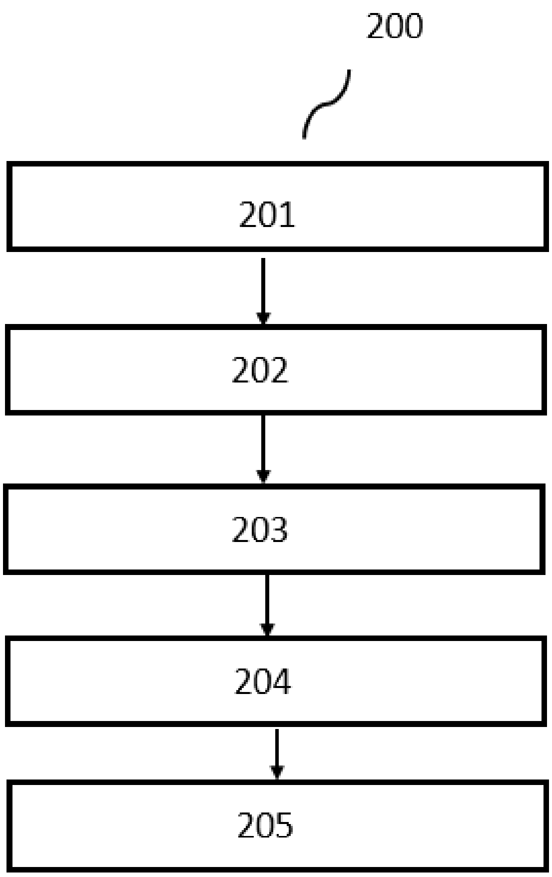

Figure 2

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

3A

3B

Figure 3

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