rotating, or cropping the image, which generalize the developed model with respect to camera properties such as focal length, axis skew, position, and rotation.

[0005] To account for these variations and deploy the trained network on various sources, training data may be augmented or manipulated to increase robustness of the trained model. These approaches, however, typically prevent models from learning effectively for any particular camera configuration by applying transformations that modify camera properties in the augmented images.

SUMMARY

[0006] According to a first aspect, there is provided a method for training a set of parameters of a predictive computer model for use by a model application system, the method including: identifying a set of images captured by a first plurality of cameras which are affixed to one or more image collection systems, wherein the one or more image collection systems and the model application system have the same camera configuration; for each image in the set of images, identifying a training output for the image; for one or more images in the set of images, generating an augmented image for a set of augmented images by: generating an augmented image for a set of augmented images by modifying the image with an image manipulation function that maintains camera properties of the image, such that angle, scale, and/or pose associated with the image is preserved, and associating the augmented training image with the training output of the image; training the set of parameters of the predictive computer model to predict the training output based on an image training set including the images and the set of augmented images, wherein the trained predictive computer model is configured to predict a presence of objects in input images for use in autonomous or semi-autonomous control of the model application system, wherein a second plurality of cameras are affixed to the model application system.

[0007] According to a second aspect, there is provided a system comprising one or more processors and non-transitory computer storage media storing instructions that when executed by the one or more processors, cause the processors to perform operations including: identifying a set

of images captured by a first plurality of cameras which are affixed to one or more image collection systems; for each image in the set of images, identifying a training output for the image; for one or more images in the set of images, generating an augmented image for a set of augmented images by: generating an augmented image for a set of augmented images by modifying the image with an image manipulation function that maintains camera properties of the image, such that angle, scale, and/or pose associated with the image is preserved, and associating the augmented training image with the training output of the image; training a predictive computer model to predict the training output based on an image training set including the images and the set of augmented images, wherein the trained predictive computer model is configured to predict a presence of objects in input images for use in autonomous or semi-autonomous control of a model application system having the same camera configuration as the one or more image collection systems, wherein a second plurality of cameras are affixed to the model application system.

[0008] According to a third aspect, there is provided a non-transitory computer-readable medium having instructions for execution by a processor, the instructions when executed by the processor causing the processor to: identify a set of images captured by a first plurality of cameras which are affixed to one or more image collection systems; for each image in the set of images, identify a training output for the image; for one or more images in the set of images, generate an augmented image for a set of augmented images by: generate an augmented image for a set of augmented images by modifying the image with an image manipulation function that maintains camera properties of the image, such that angle, scale, and/or pose associated with the image is preserved, and associate the augmented training image with the training output of the image; train a predictive computer model to predict the training output based on an image training set including the images and the set of augmented images, wherein the trained predictive computer model is configured to predict a presence of objects in input images for use in autonomous or semi-autonomous control of a model application system having the same camera configuration as the one or more image collection systems, wherein a second plurality of cameras are affixed to the model application

system, and wherein an individual configuration indicates a same position and/or orientation with respect to the one or more image collection systems and the model application system, and wherein consistent artifacts and effects due to the individual configuration are included in images captured by a first camera of the first plurality of cameras which is affixed according to the individual configuration and a second camera of the second plurality of cameras which is affixed according to the individual configuration.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Fig. 1 is block diagram of an environment for computer model training and deployment according to one embodiment.

[0010] Fig. 2 illustrates example images captured with the same camera characteristics.

[0011] Fig. 3 is a block diagram of components of a model training system, according to one embodiment.

[0012] Fig. 4 is a data flow diagram showing an example of generating augmented images based on a labeled training image, according to one embodiment.

[0013] The figures depict various embodiments of the present invention for purposes of illustration only. One skilled in the art will readily recognize from the following discussion that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles of the invention described herein.

DETAILED DESCRIPTION

[0014] One embodiment is a system that trains a computer model with images which have been augmented to maintain the camera properties of the originally-captured images. These camera properties may include intrinsic or extrinsic properties of the camera. Such intrinsic properties may include characteristics of the sensor itself, such as dynamic range, field of view, focal length, and lens distortion. Extrinsic properties may describe the configuration of the camera with respect to the captured environment, such as the angle, scale, or pose of a camera.

CLAIMS

1. A method for training a set of parameters of a predictive computer model for use by a model application system, the method comprising:

identifying a set of images captured by a first plurality of cameras which are affixed to one or more image collection systems, wherein the one or more image collection systems and the model application system have the same camera configuration;

for each image in the set of images, identifying a training output for the image;

for one or more images in the set of images, generating an augmented image for a set of augmented images by:

generating an augmented image for a set of augmented images by modifying the image with an image manipulation function that maintains camera properties of the image, such that angle, scale, and/or pose associated with the image is preserved, and

associating the augmented training image with the training output of the image;

training the set of parameters of the predictive computer model to predict the training output based on an image training set including the images and the set of augmented images,

wherein the trained predictive computer model is configured to predict a presence of objects in input images for use in autonomous or semi-autonomous control of the model application system, wherein a second plurality of cameras are affixed to the model application system.

2. The method of claim 1, wherein an individual configuration indicates a same position and/or orientation with respect to the one or more image collection systems and the model application system, and wherein consistent artifacts and effects due to the individual configuration are included in images captured by a first camera of the first plurality of cameras which is affixed according to the individual configuration and a second camera of the second plurality of cameras which is affixed according to the individual configuration.

3. The method of claim 1, wherein the image training set does not include images generated by image manipulation functions that modify camera properties of an image.

4. The method of claim 3, wherein the image manipulation functions that modify camera properties include crop, pad, horizontal or vertical flip, or affine transformations.

5. The method of claim 1, wherein the image manipulation function is cutout, hue, saturations, value jitter, salt and pepper, domain transfer or any combination thereof.

6. The method of claim 1, wherein the image manipulation function is a cutout applied to the image, wherein a location of the cutout is selected based on an expected view of the first plurality of cameras with respect to the one or more image collection systems, and wherein the location includes a center of the image which depicts a direction of travel.

7. The method of claim 1, wherein the image manipulation function is a cutout applied to a portion of the image, wherein a location of the cutout is selected based on an expected view of a first camera, of the first plurality of cameras, with respect to the one or more image collection systems, and wherein the location includes artifacts which are always present in images captured by the first camera.

8. The method of claim 1, wherein the image manipulation function is a cutout applied to a portion of the image that partially overlaps with the location of the training output in the image, the cutout being a solid color and the cutout being blended with regions near the cutout.

9. A system comprising one or more processors and non-transitory computer storage media storing instructions that when executed by the one or more processors, cause the processors to perform operations comprising:

identifying a set of images captured by a first plurality of cameras which are affixed to one or more image collection systems;

for each image in the set of images, identifying a training output for the image;

for one or more images in the set of images, generating an augmented image for a set of augmented images by:

generating an augmented image for a set of augmented images by modifying the image with an image manipulation function that maintains camera properties of the image, such that angle, scale, and/or pose associated with the image is preserved, and

associating the augmented training image with the training output of the image;

training a predictive computer model to predict the training output based on an image training set including the images and the set of augmented images,

wherein the trained predictive computer model is configured to predict a presence of objects in input images for use in autonomous or semi-autonomous control of a model application system having the same camera configuration as the one or more image collection systems, wherein a second plurality of cameras are affixed to the model application system.

10. The system of claim 9, wherein an individual configuration indicates a same position and/or orientation with respect to the one or more image collection systems and the model application system, and wherein consistent artifacts and effects due to the individual configuration are included in images captured by a first camera of the first plurality of cameras which is affixed according to the individual configuration and a second camera of the second plurality of cameras which is affixed according to the individual configuration.

11. The system of claim 9, wherein the image training set does not include images generated by image manipulation functions that modify camera properties of an image, and wherein the image manipulation functions that modify camera properties include crop, pad, horizontal or vertical flip, or affine transformations.

12. The system of claim 9, wherein the image manipulation function is cutout, hue, saturations, value jitter, salt and pepper, domain transfer or any combination thereof.

13. The system of claim 9, wherein the image manipulation function is a cutout applied to a portion of the image, wherein a location of the cutout is selected based on an expected view of a first camera, of the first plurality of cameras, with respect to the one or more image collection systems, and wherein the location includes artifacts which are always present in images captured by the first camera.

14. A non-transitory computer-readable medium having instructions for execution by a processor, the instructions when executed by the processor causing the processor to:

identify a set of images captured by a first plurality of cameras which are affixed to one or more image collection systems;

for each image in the set of images, identify a training output for the image;

for one or more images in the set of images, generate an augmented image for a set of

augmented images by:

generate an augmented image for a set of augmented images by modifying the image with an image manipulation function that maintains camera properties of the image, such that angle, scale, and/or pose associated with the image is preserved, and

associate the augmented training image with the training output of the image;

train a predictive computer model to predict the training output based on an image training set including the images and the set of augmented images,

wherein the trained predictive computer model is configured to predict a presence of objects in input images for use in autonomous or semi-autonomous control of a model application system having the same camera configuration as the one or more image collection systems, wherein a second plurality of cameras are affixed to the model application system, and

wherein an individual configuration indicates a same position and/or orientation with respect to the one or more image collection systems and the model application system, and wherein consistent artifacts and effects due to the individual configuration are included in images captured by a first camera of the first plurality of cameras which is affixed according to the individual configuration and a second camera of the second plurality of cameras which is affixed according to the individual configuration.

15. The non-transitory computer-readable medium of claim 14, wherein the image training set does not include images generated by image manipulation functions that modify camera properties of an image.

16. The non-transitory computer-readable medium of claim 15, wherein the image manipulation functions that modify camera properties include crop, pad, horizontal or vertical clip, or affine transformations.

17. The non-transitory computer-readable medium of claim 14, wherein the image manipulation function is cutout, hue, saturations, value jitter, salt and pepper, domain transfer or any combination thereof.

18. The non-transitory computer-readable medium of claim 14, wherein the image manipulation function is a cutout applied to a portion of the image, wherein a location of the cutout

is selected based on an expected view of the first camera, of the first plurality of cameras, with respect to the one or more image collection systems, and wherein the location includes artifacts which are always present in images captured by the first camera.

19. The non-transitory computer-readable medium of claim 14, wherein the image manipulation function is a cutout applied to a region that partially overlaps with the location of the training output in the image, the cutout being a solid color and the cutout being blended with regions near the cutout.

20. The non-transitory computer-readable medium of claim 14, wherein the image manipulation function is a cutout applied to a portion of the image, wherein a location of the cutout is selected based on an expected view of the first camera, of the first plurality of cameras, with respect to the one or more image collection systems, and wherein the location includes a center of the image which depicts a direction of travel.