

CLAIMS

1. A computer-implemented method comprising:
 - receiving training data representing a plurality of images at a deep neural network training system,
 - the plurality of images depicting objects having an object sub-type, and
 - the deep neural network training system to train a deep neural network to recognize the object sub-type for one of the objects depicted in the plurality of images, wherein multiple object sub-types for which corresponding objects are depicted in the plurality of images each correspond to the same object type;
 - for a plurality of iterations:
 - selecting particular training data for an image from the plurality of images;
 - determining whether to randomly permute a value of a property of the selected image that is represented by the particular training data;
 - providing, to the deep neural network, the particular training data for the selected image or the randomly permuted particular training data for the selected image;
 - receiving, from the deep neural network, output data indicating:
 - a predicted label for a particular object sub-type for an object depicted in the selected image, and
 - a confidence score that represents a likelihood that the object depicted in the selected image has the particular object sub-type; and
 - updating one or more weights in the deep neural network using:
 - an expected output value that indicates an actual object sub-type label for the object depicted in the selected image,
 - the predicted label for the particular object sub-type, and
 - the confidence score that represents the likelihood that the object depicted in the selected image has the particular object sub-type; and
 - providing the deep neural network to a mobile device for use detecting whether one or more images depict objects having the particular object sub-type.

2. The method of claim 1, wherein updating the one or more weights in the deep neural network comprises updating the one or more weights in the deep neural network using, for each of two or more images from the plurality of images, a respective expected output value, the respective predicted label for the particular object sub-type, and the respective confidence score.
3. The method of claim 1, wherein:
 - receiving, from the deep neural network, output data indicating the predicted label for the particular object sub-type for the object depicted in the selected image and the confidence score that represents the likelihood that the object depicted in the selected image has the particular object sub-type comprises:
 - receiving, from the deep neural network, the output data that indicates a predicted location of the object in the selected image; and
 - updating the one or more weights in the deep neural network comprises:
 - updating the one or more weights in the deep neural network using the expected output value, the predicted label for the particular object sub-type, the respective confidence score, an actual location of the object in the selected image, and the predicted location of the object in the selected image.
4. The method of claim 1, wherein determining whether to randomly permute a value of a property of the selected image that is represented by the particular training data comprises determining whether to randomly perform one or more of the following:
 - crop a portion of the selected image;
 - adjust a zoom level for at least a portion of the selected image;
 - adjust a stretch value for at least a portion of the selected image;
 - transpose the selected image over an axis;
 - adjust at least one color depicted in the selected image;
 - adjust an alpha jitter for at least a portion of the selected image;
 - adjust a shear distortion for at least a portion of the selected image; or
 - adjust a homography for at least a portion of the selected image.

5. The method of claim 4, wherein:
- each iteration comprises selecting two or more images from the plurality of images;
 - and
 - determining whether to randomly permute a value of a property of the selected image that is represented by the particular training data comprises:
 - selecting, for each iteration, one or more particular images from the selected two or more images; and
 - determining, for each of the one or more particular images, a property to randomly permute.
6. A non-transitory computer storage medium encoded with instructions that, when executed by one or more computers, cause the one or more computers to perform operations comprising:
- receiving data for an image that depicts a plurality of objects;
 - determining whether to create multiple overlapping portions of data for the image,
 - the multiple overlapping portions each including a proper subset of data for the image;
 - in response to determining to create the multiple overlapping portions of data for the image:
 - creating the multiple overlapping portions of data for the image using the received data for the image;
 - providing, to a deep neural network, data for each of the multiple overlapping portions separately to cause the deep neural network to predict object sub-type labels for each of the plurality of objects;
 - receiving, from the deep neural network for some of the objects in the plurality of objects:
 - output data that represents a sub-type label that indicates a likely object sub-type for the object, and
 - an object location that indicates a location of the object in the image; and
 - generating, using the output data for some of the objects in the plurality of objects, a representation of the image that indicates, for some of the objects in the plurality of objects, a location of the object in the image and the sub-type label that indicates the likely object sub-

type for the object.

7. The computer storage medium of claim 6, the operations comprising:
receiving, for some of the objects in the plurality of objects, a corresponding confidence score that indicates a likelihood that the object is of the object sub-type, wherein:
determining whether to create multiple overlapping portions of data for the image that each include a proper subset of data for the image comprises determining, for each of the confidence scores, whether the corresponding confidence score satisfies a threshold confidence score; and

creating the multiple overlapping portions of data for the image is responsive to determining that at least one of the confidence scores does not satisfy the threshold confidence score.

8. The computer storage medium of claim 6, wherein:
determining whether to create multiple overlapping portions of data for the image that each include a proper subset of data for the image comprises determining whether a current processing duration satisfies a threshold processing duration; and
creating the multiple overlapping portions of data for the image is responsive to determining that the current processing duration satisfies the threshold processing duration.

9. The computer storage medium of claim 6, the operations comprising:
receiving, for some of the objects in the plurality of objects, a corresponding confidence score that indicates a likelihood that the object is of a corresponding object sub-type; and
determining, for at least some of the objects using the corresponding confidence scores, whether to use a disambiguation deep neural network, different from the deep neural network, to determine an object sub-type label for the object.

10. The computer storage medium of claim 9, wherein:
receiving, for some of the objects in the plurality of objects, a corresponding confidence score that indicates a likelihood that the object is of a corresponding object sub-type comprises receiving, for some of the objects in the plurality of objects, multiple corresponding confidence scores each of which indicates a likelihood that the object is of an

object sub-type for which the deep neural network is trained; and

determining, for at least some of the objects using the corresponding confidence scores, whether to use a disambiguation deep neural network, different from the deep neural network, to determine an object sub-type label for the object comprises determining, for at least some of the objects, whether two of the corresponding confidence scores each satisfy a threshold confidence score.

11. The computer storage medium of claim 9, wherein:

receiving, for some of the objects in the plurality of objects, a corresponding confidence score that indicates a likelihood that the object is of a corresponding object sub-type comprises receiving, for some of the objects in the plurality of objects, multiple corresponding confidence scores each of which indicates a likelihood that the object is of an object sub-type for which the deep neural network is trained; and

determining, for at least some of the objects using the corresponding confidence scores, whether to use a disambiguation deep neural network, different from the deep neural network, to determine an object sub-type label for the object comprises determining, for at least some of the objects, whether two of the corresponding confidence scores are within a threshold distance from each other.

12. The computer storage medium of claim 6, wherein:

receiving, from the deep neural network for some of the objects in the plurality of objects, the output data that represents the sub-type label that indicates a likely object sub-type for the object, and the object location that indicates a location of the object in the image comprises receiving, for some of the objects in the plurality of objects, a corresponding confidence score that indicates a likelihood that the object is of the object sub-type; and

generating, using the output data for some of the objects in the plurality of objects, the representation of the image that indicates, for some of the objects in the plurality of objects, the location of the object in the image and the sub-type label that indicates the likely object sub-type for the object comprises:

determining that at least one of the confidence scores does not satisfy a threshold confidence score;

determining that a current processing time exceeds a threshold

processing time; and

in response to determining that a current processing time exceeds a threshold processing time, determining to generate the representation of the image despite determining that at least one of the confidence scores does not satisfy a threshold confidence score.

13. The computer storage medium of claim 6, the operations comprising:

determining, using the representation of the image, a stock of one or more objects from the plurality of objects; and

determining, using the stock of the one or more objects, a quantity of objects to request for delivery to a physical location that includes an area depicted in the image.

14. The computer storage medium of claim 6, wherein receiving the data for the image that depicts the plurality of objects comprises:

receiving, for each of two or more images, first data for the respective image;

determining, using the first data for each of the two or more images, that the two or more images depict the same physical content; and

in response to determining that the two or more images depict the same physical content determining, using the first data for each of the two or more images, the data for the image that depicts the plurality of objects by stitching together the two or more images.

15. A deep neural network training system comprising one or more computers and one or more storage devices on which are stored instructions that are operable, when executed by the one or more computers, to cause the one or more computers to perform operations comprising:

receiving training data representing a plurality of images,

the plurality of images depicting objects having an object sub-type, and

the deep neural network training system training a deep neural network to recognize the object sub-type for one of the objects depicted in the plurality of images,

wherein multiple object sub-types for which corresponding objects are depicted in the plurality of images each correspond to the same object type;

for a plurality of iterations:

selecting particular training data for an image from the plurality of images;
 determining whether to randomly permute a property of the selected image that
 is represented by the particular training data;
 providing, to the deep neural network, the particular training data for the
 selected image or the randomly permuted particular training data for the selected image;
 receiving, from the deep neural network, output data indicating:
 a predicted label for a particular object sub-type for an object depicted
 in the selected image,
 location information defining a region of the selected image in which
 the object is likely depicted, and
 a confidence score that represents a likelihood that the object exists in
 the region of the selected image defined by the location information; and
 updating one or more weights in the deep neural network using:
 an expected output value that indicates an actual object sub-type label
 for the object depicted in the selected image,
 the predicted label for the particular object sub-type, and
 the confidence score that represents the likelihood that the object exists
 in the region of the selected image defined by the location information; and
 providing the deep neural network to a mobile device for use detecting whether one or
 more images depict objects having the particular object sub-type.

16. The system of claim 15, wherein updating the one or more weights in the deep neural
 network comprises updating the one or more weights in the deep neural network using, for
 each of two or more images from the plurality of images, a respective expected output value,
 the respective predicted label for the particular object sub-type, and the respective confidence
 score.

17. The system of claim 15, wherein:
 the confidence score comprises an object confidence score;
 receiving, from the deep neural network, output data indicating the predicted label for
 the particular object sub-type for the object depicted in the selected image and the confidence
 score that represents the likelihood that the object exists in the region of the selected image

defined by the location information comprises:

receiving, from the deep neural network, the output data that indicates a label confidence score that represents a likelihood that the object depicted in the selected image has the particular object sub-type; and

updating the one or more weights in the deep neural network comprises:

updating the one or more weights in the deep neural network using the expected output value, the predicted label for the particular object sub-type, the respective label confidence score, the respective object confidence score, an actual location of the object in the selected image, and the location information defining a region of the selected image in which the object is likely depicted.

18. The system of claim 15, wherein determining whether to randomly permute a value of a property of the selected image that is represented by the particular training data comprises determining whether to randomly perform one or more of the following:

crop a portion of the selected image;
adjust a zoom level for at least a portion of the selected image;
adjust a stretch value for at least a portion of the selected image;
transpose the selected image over an axis;
adjust at least one color depicted in the selected image;
adjust an alpha jitter for at least a portion of the selected image;
adjust a shear distortion for at least a portion of the selected image; or
adjust a homography for at least a portion of the selected image.

19. The system of claim 18, wherein:

each iteration comprises selecting two or more images from the plurality of images;
and

determining whether to randomly permute a value of a property of the selected image that is represented by the particular training data comprises:

selecting, for each iteration, one or more particular images from the selected two or more images; and

determining, for each of the one or more particular images, a property to randomly permute.

20. A non-transitory computer storage medium encoded with instructions that, when executed by one or more computers, cause the one or more computers to perform operations comprising:

receiving data for a particular image that depicts a plurality of objects;

providing, to a deep neural network trained to detect objects of a particular type using one or more anchor boxes that each include location data that identifies a respective box in an image, the data to cause the deep neural network to predict object sub-type labels for each of the plurality of objects, the object sub-type labels for objects having a sub-type of the particular type;

receiving, from the deep neural network for some of the objects in the plurality of objects, output data a) determined by the deep neural network using the one or more anchor boxes applied to the particular image and b) that represents a sub-type label that indicates a likely object sub-type for the object, and an object location that indicates a location of the object in the particular image; and

generating, by the one or more computers and using the output data for some of the objects in the plurality of objects, instructions for presentation of a user interface that includes a representation of the particular image that indicates, for some of the objects in the plurality of objects, a location of the object in the particular image and the sub-type label that indicates the likely object sub-type for the object; and

providing the instructions to a display to cause the display to present the user interface.