



(51) International Patent Classification:

G06V 30/412 (2022.01) G06V 30/414 (2022.01)
G06V 30/19 (2022.01) G06F 21/62 (2013.01)

(21) International Application Number:

PCT/US2024/033282

(22) International Filing Date:

10 June 2024 (10.06.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

18/214,108 26 June 2023 (26.06.2023) US

(71) Applicant: **AMAZON TECHNOLOGIES, INC.**
[US/US]; PO BOX 81226, Seattle, Washington 98108-1226 (US).

(72) Inventors: **LI, Kun**; 410 Terry Avenue North, Seattle, Washington 98109-5210 (US). **MISHRA, Pragyana K.**;

410 Terry Avenue North, Seattle, Washington 98109-5210 (US).

(74) Agent: **MATSUBAYASHI, Hugh H.** et al.; 925 4th Ave., Suite 2900, Seattle, Washington 98104-1158 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV,

(54) Title: PRIVACY-PRESERVING TRAINING AND EVALUATION OF COMPUTER VISION MODELS

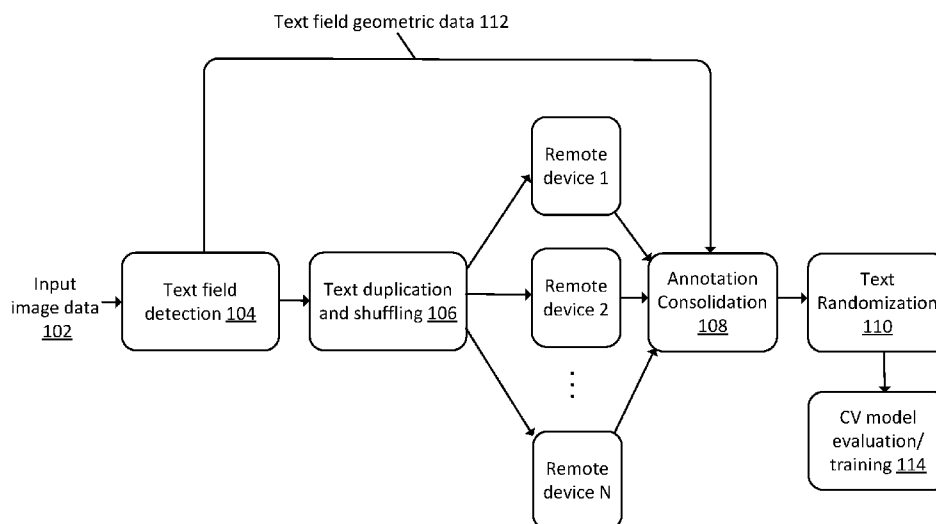


FIG. 1

(57) Abstract: Devices and techniques are generally described for privacy preservation for computer vision models. In some examples, a first field of text and a second field of text may be detected in a first image. A first alpha-numeric text string may be detected in the first field and a second alpha-numeric text string may be detected in the second field. A first sub-image including the first alpha-numeric text string may be generated and a second sub-image including the second alpha-numeric text string may be generated. The first sub-image may be sent to a first computing device for annotation and the second sub-image may be sent to a second computing device for annotation.

GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

PRIVACY-PRESERVING TRAINING AND EVALUATION OF COMPUTER VISION MODELS

BACKGROUND

[0001] In order to comply with various government regulations and best practices, stewards of data are required to maintain strict control over the usage, distribution, handling, and retention of personal data related to individuals. In various examples, this includes instituting capabilities to retrieve and present all personal data on demand, delete all personal data on demand, and adhere to complicated time- and rules-based retention and deletion schedules for personal data.

BRIEF DESCRIPTION OF DRAWINGS

[0002] FIGS. 1 is a flow diagram illustrating an example system that may be used to preserve privacy during training and/or evaluation of computer vision models, in accordance with various aspects of the present disclosure.

[0003] FIG. 2A illustrates an example of raw image data including potentially sensitive and/or personally identifiable information, in accordance with various examples described herein.

[0004] FIG. 2B illustrates the example image data of FIG. 2A including text field detection, in accordance with various examples described herein.

[0005] FIG. 2C illustrates sub-images of alpha-numeric text strings generated for each text field detected in FIG. 2B and a background image with detected text removed, in accordance with various examples described herein.

[0006] FIGS. 2D and 2E illustrate example annotation interfaces that may be used in accordance with the examples depicted in FIGS. 2A-2C, in accordance with various aspects of the present disclosure.

[0007] FIG. 2F illustrates an example field attribute type annotation interface image with randomized alphanumeric strings, in accordance with various aspects of the present disclosure.

[0008] FIG. 3 is a block diagram showing an example architecture of a computing device that may be used in accordance with various embodiments described herein.

[0009] FIG. 4 depicts an example process for privacy preservation for computer vision model training and/or evaluation, in accordance with various aspects of the present disclosure.

DETAILED DESCRIPTION

[0010] In the following description, reference is made to the accompanying drawings that illustrate several examples of the present invention. It is understood that other examples may be

utilized and various operational changes may be made without departing from the scope of the present disclosure. The following detailed description is not to be taken in a limiting sense, and the scope of the embodiments of the present invention is defined only by the claims of the issued patent.

[0011] Storage and/or use of data related to a particular person or entity (e.g., personally identifiable information and/or other sensitive data) may be required to comply with regulations, privacy policies, and/or legal requirements of the relevant jurisdictions. In many cases, users may be provided with the option of opting out of storage and/or usage of personal data and/or may select particular types of personal data that may be stored while preventing aggregation and storage of other types of personal data. Additionally, aggregation, storage, and/or use of personal data may be compliant with privacy controls, even if not legally subject to them. For example, storage and/or use of personal data may be subject to acts and regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation (GDPR), and/or other data privacy frameworks.

[0012] Maintaining the integrity and confidentiality of sensitive data (e.g., health data related to an individual, sensitive financial information, etc.) involves employment of specific measures against accidental loss, and unauthorized or unlawful processing of such data. Personally Identifiable Information (PII) is data relating directly or indirectly to an individual, from which the identity of the individual can be determined. Examples of PII include patient names, addresses, phone numbers, Social Security numbers, bank account numbers, etc. Images and videos of insurance cards, driver licenses, prescriptions, pill bottle labels, passports, and medical bills typically include PII.

[0013] This makes it very difficult to develop, test, and deploy machine learning algorithms that detect and extract names, address, phone numbers, and other PII from image data (e.g., single images, video frames, and videos). For example, although existing computer vision algorithms are easily able to detect various text and fields in an image (e.g., an image of a driver's license) and use such information to auto-populate a user's account (assuming all applicable user permissions), the data used to develop, test, and deploy such machine learning models have to be stored, transmitted, and processed in a secure manner. As such, it may be cumbersome to deploy such models in practice as numerous security protocols may need to be adhered to in order to preserve user privacy and/or to prevent improper usage and/or storage of sensitive data.

[0014] In various cases, such computer vision algorithms are tested and validated for performance in secure environments—which can often make the process cumbersome, time consuming, restrictive, and complex. For example, the processing pipelines, libraries, services,

storage, and/or network handling for the data may add to security risks. Described herein are systems and techniques that may be used to process, store, and transfer the image data (which may include sensitive data such as PII) while maintaining privacy and security of such data. In addition the techniques described herein are able to train, test, and deploy such computer vision models using techniques that do not add to the overall cost or latency of model training or deployment relative to conventional approaches which may not offer such data security.

[0015] In various examples, object detectors are computer vision machine learning models that are able locate and/or classify objects detected in frames of image data. Typically, the output of an object detector model is a “bounding box” or “region of interest” surrounding a group of pixels and a label (label data) classifying that bounding box or region of interest to a particular class for which the object detector has been trained. For example, an object detector may be trained to classify dogs and cats. If an input image includes first pixels representing a dog and second pixels representing a cat, the object detector may output two bounding boxes (e.g., output bounding box data). The first bounding box may surround the first pixels and may be labeled as “dog.” Similarly, the second bounding box may surround the second pixels and may be labeled as “cat.” In some other examples, an object detector may be trained to detect text present in an image. The object detector may provide bounding boxes or may otherwise identify regions in the image in which text is detected.

[0016] Bounding boxes may be of any shape. For example, bounding boxes are rectangular and may be defined by the four pixels addresses that correspond to the corners of the bounding box. In some examples, bounding boxes are defined by a perimeter of pixels surrounding pixels predicted to correspond to some object which the object detector has been trained to detect. More generally, object detectors may detect regions of interest (RoIs). Bounding boxes are one example of an RoI that may be detected by an object detector. However, RoIs may be defined in other ways apart from bounding boxes. For example, pixels corresponding to a detected object may be classified to distinguish these pixels from those pixels representing other objects. An object detector may be implemented using a convolutional neural network (CNN), a vision transformer, etc.

[0017] For example, in order to generate a training dataset that is able to automatically process user health insurance information, it may not be possible to provide an image of a user's insurance card to an annotator and/or annotation system to have the relevant text data (e.g., Insurance Policy Number, User name, Group Number, etc.) annotated for privacy and security reasons. Described herein are various systems and techniques for privacy-preserving training and evaluation of computer vision models. In various examples, the approaches may involve fragmentation of an image into a plurality of sub-images so that any information in a smaller

image fragment (sub-image) cannot be used to identify or determine any sensitive information about a person to which the overall image pertains. For example, an image with potential PII information is fragmented or cut into random sub-images. Each sub-image may represent a smaller segment of the entire image thereby ensuring that any PII information cannot be recovered from the smaller sub-image. The front end processing of an image involves detection of contiguous text and breaking up the image such that the contiguous text is fragmented—thereby guaranteeing privacy and security. These fragmented sub-images may be annotated and/or otherwise processed using normal annotation/processing pipelines without loss of security/privacy. After annotation, the sub-images may be consolidated to reassemble the original image using geometrical information describing the position/orientations/locations of the fragments within the original image. Such annotated images may be used to develop (e.g., train) and deploy image processing algorithms such as optical character recognition (OCR), segmentation, detection, classification of objects without the need to harden the systems to maintain security.

[0018] Machine learning techniques, such as those described herein, are often used to form predictions, solve problems, recognize objects in image data for classification, etc. For example, in the context of object detection/classification, a machine learning architecture may learn to analyze input images, detect various object classes (e.g., cats, dogs, text fields, etc.), and/or distinguish between instances of objects appearing in the images. In various examples, machine learning models may perform better than rule-based systems and may be more adaptable as machine learning models may be improved over time by retraining the models as more and more data becomes available. Accordingly, machine learning techniques are often adaptive to changing conditions. Deep learning algorithms, such as neural networks, are often used to detect patterns in data and/or perform tasks.

[0019] Generally, in machine learned models, such as neural networks, parameters (weights) control activations in neurons (or nodes) within layers of the machine learned models. The weighted sum of activations of each neuron in a preceding layer may be input to an activation function (e.g., a sigmoid function, a rectified linear units (ReLU) function, etc.). The result determines the activation of a neuron in a subsequent layer. In addition, a bias value can be used to shift the output of the activation function to the left or right on the x-axis and thus may bias a neuron toward activation.

[0020] Generally, in machine learning models, such as neural networks, after initialization, annotated training data may be used to generate a cost or “loss” function that describes the difference between expected output of the machine learning model and actual output. The parameters (e.g., weights and/or biases) of the machine learning model may be updated to

minimize (or maximize) the cost. For example, the machine learning model may use a gradient descent (or ascent) algorithm to incrementally adjust the weights to cause the most rapid decrease (or increase) to the output of the loss function. The method of updating the parameters of the machine learning model is often referred to as back propagation.

[0021] FIGS. 1 is a flow diagram illustrating an example system that may be used to preserve privacy during training and/or evaluation of computer vision models, in accordance with various aspects of the present disclosure. Input image data 102 may be an image that may include various text (including potentially sensitive data, such as PII). The input image data 102 may be one or more frames of a video, a single image, or multiple images.

[0022] Text field detection 104 may involve detecting contiguous text within the input image data 102. In various examples, individual text fields may be detected for each grouping of contiguous alpha-numeric characters (e.g., characters without a space). Accordingly, "12345" in an image may be detected as a single text field, while "John Smith" may be detected as two contiguous text fields (one field for "John" and another field for "Smith"). Other techniques to fragment detected text fields may be used, depending on the desired implementation. For example, contiguous text may be broken up into fragments of four (or any other desired number) or fewer contiguous characters. In various examples the alpha-numeric text of a given field may be fragmented into fragments that include less than a total amount of the text detected in the field.

[0023] The text field detection 104 may be performed using a pre-trained optical character recognition (OCR) component, an object detector trained to detect text, etc. Text field geometric data 112 may include information describing the location (e.g., within the image frame), the skew, and/or the three-dimensional rotation (e.g., the angle of rotation of text shown in an image with respect to one or more axes), of each text field detected in the image. As described in further detail below, the text field geometric data 112 may be used to recreate the input image data 102 from fragments of the input image data 102 made up of sub-images of each detected text field and the background (non-text portions) of the original input image data 102. In various examples, an affine transformation may be performed on the sub-images of the detected text fields and the background in order to generate a new two dimensional image version of the original input image data 102 (e.g., to correct for a poor camera angle used to capture an image of a surface on which the text is printed).

[0024] Text duplication and shuffling 106 includes generation of a sub-image of the text of each detected text field detected at 104. Accordingly, for each detected contiguous alpha-numeric string of text a sub-image may be generated. The sub-images may be sent to various different remote devices for annotation. Annotation may include for example, verifying that text

detected by a computer vision model for the sub-image is accurate with respect to the sub-image of the text. In some other examples, annotation may include having an annotator type the alpha-numeric string shown in the sub-image. Advantageously, by sending the different sub-images to different computing devices and/or different annotators privacy is maintained as any given computing device/annotator only has access to that specific contiguous alpha-numeric string without any other context that may be used to ascertain PII or sensitive information.

[0025] Annotation consolidation 108 may receive the annotated text fields from the remote devices 1, 2, ..., N and the text field geometric data 112. In addition, the annotation consolidation 108 may generate and/or receive a background image of the input image data 102 with the detected text removed (but with locations of the various text fields detected at text field detection 104 known). In some examples, annotation consolidation 108 may store a reassembled version of the input image data 102 that includes the various detected and annotated text fields. Such re-constituted images may be used to train or re-train an object detector, an OCR model, and/or some other computer vision model.

[0026] Since the various text fields, the alpha-numeric text strings within those fields, and the respective locations of such fields are known, text randomization 110 may be used to replace the alpha-numeric text strings with randomized alpha-numeric text strings (e.g., fake information that does not divulge any PII) while maintaining the formatting of the input image. In various examples, in order to maintain the characteristics of the alpha-numeric text strings, computer-implemented logic may be used that replaces a given character with a random character of the same type. For example, capital letters of the alphabet may be replaced with random capital letters of the same alphabet, while lower-case letters of the alphabet may be replaced with random lower-case letters of the same alphabet. For example, text randomization may replace the alpha-numeric text string "A2z4" with the alpha-numeric text string "L1b7" and may replace the alpha-numeric text string "Susan" with the alpha-numeric text string "Bagrf." In this way, alpha-numeric text strings for particular fields maintain the same length (in terms of numbers of characters) and other characteristics without divulging sensitive and/or personally-identifiable information. Once the text has been randomized in this way, the image that includes randomized alpha-numeric text in the detected text fields may be sent to an annotator that may annotate the text fields with their corresponding attribute type (e.g., a field that includes a randomized user's name may be identified as a "Name" field, while a field that includes a randomized account number may be identified as an "Account Number" field). These techniques are described in further detail below.

[0027] Upon receiving all relevant annotations (e.g., of the alpha-numeric text itself and/or of the attribute types for the various fields detected in the input image data 102) the annotated

data may be incorporated into a training data set and used to train/re-train the relevant computer vision model(s) at block 114. In various other examples, the annotations may be used to evaluate the performance of the relevant computer-vision model(s). For example, the accuracy/precision/recall of a computer vision model may be evaluated based on its ability to correctly detect text or detect a particular type of text field from input image data.

[0028] FIG. 2A illustrates an example of raw image data including potentially sensitive and/or personally identifiable information, in accordance with various examples described herein. The example image in FIG. 2A is an insurance card that may include PII such as the card-holder's name, ID number, etc. It may be impermissible to directly send such an image to an annotator and/or annotator's may be required to be specifically vetted, trained, and/or evaluated in order to handle such sensitive data.

[0029] FIG. 2B illustrates the example image data of FIG. 2A including text field detection, in accordance with various examples described herein. In the example of FIG. 2B, text field detection 104 may have been performed using a pre-trained optical character recognition (OCR) model, an object detector trained to detect text, etc. As shown, individual text fields have been detected for each grouping of contiguous alpha-numeric characters (e.g., characters without a space). The detections are represented using respective bounding boxes in FIG. 2B.

Accordingly, the identification number "JQP123X45678" has been detected as a single text field, while "John Q" has been detected as two contiguous text fields (one field for "John" and another field for "Q"). Other techniques to fragment detected text fields may be used, depending on the desired implementation. For example, contiguous text may be broken up into fragments of four (or any other desired number) or fewer contiguous characters.

[0030] FIG. 2C illustrates sub-images of alpha-numeric text strings generated for each text field detected in FIG. 2B and a background image with detected text removed, in accordance with various examples described herein. For each detected text field in FIG. 2B, a sub-image may be generated representing the alpha-numeric text in that field. For example, for text field 250, a sub-image 252 representing the alpha-numeric text string "INSURCO" has been generated. Similarly, for text field 254, a sub-image 256 representing the alpha-numeric text string "HEALTHSHIELD" has been generated. Background image 212 may also be generated which may include blank versions of the detected text fields. The blank detected fields may be represented as the text field geometric data 112 from FIG. 1. In some examples, the background image 212 may be sent to an annotator for annotation of any undetected text field which remained in the background image 212. Such annotations may be used to improve the text field detection 104 (e.g., by retraining the object detector, OCR model, or other text detector used to detect the text in the input image).

[0031] FIGS. 2D and 2E illustrate example annotation interfaces that may be used in accordance with the examples depicted in FIGS. 2A-2C, in accordance with various aspects of the present disclosure. The various sub-images of alpha-numeric text strings generated for each text field detected in FIG. 2B may be sent to various different remote computing devices for annotation. Since any given computing device/annotator will receive only the sub-image of a text fragment without receiving any larger context from the overall input image data, privacy of the underlying data is preserved. In the example in FIG. 2D, the sub-image of the text is presented (e.g., the blurred text on the left-hand side detected from a blurry user-captured image) alongside an OCR model's prediction of the alpha-numeric string represented in that sub-image (the text on the right-hand side). For each such pairing of a sub-image together with the predicted text represented in that sub-image, the annotator may be provided with an interface in which the annotator may indicate whether there is a match between the alpha-numeric string in the sub-image and the alpha-numeric string detected by the OCR model (or other text detection model). This information may be used to evaluate the model performance of the OCR or other text detection model.

[0032] In the example interface in FIG. 2E, the annotator is presented with the sub-image of the alpha-numeric string and is asked to type the alpha-numeric string represented in the sub-image in the provided field in the graphical user interface. This annotation may be used as a ground truth label to train the OCR model or other text recognition model (e.g., using supervised machine learning techniques).

[0033] FIG. 2F illustrates an example field attribute type annotation interface image with randomized alphanumeric strings, in accordance with various aspects of the present disclosure. After detecting the relevant text fields and removing the alpha-numeric text strings to generate the background image 212, the text fields may be replaced with random (or pseudo-random) alpha-numeric text strings as described above in reference to text randomization 110. In the example of FIG. 2F, the alpha-numeric text of various text fields has been replaced with random alpha-numeric text. For example, the identification number from FIG. 2A (JQP123X45678) has been replaced with a pseudo-random identification number (ISS559C03936). Note that the pseudo-random identification number includes the same number of characters as the real identification number and replaces upper-case alphabet letters with random upper-case alphabet letters and numbers with random numbers. Such randomized data may be presented to an annotator along with a graphical user interface instructing the annotator to provide bounding box (or other polygonal) annotation to identify attribute types of certain fields. In the example in FIG. 2F, the annotator has been asked to identify the "Member ID" field, the "Bin_Number" field, and the "Group_ID" field and to annotate them using different patterned bounding boxes,

as shown. Because the alpha-numeric values displayed in such fields are random values, no sensitive data is divulged. Field attribute type annotation may be used to train an object detector or other computer vision to detect fields of the relevant type in input image data.

[0034] FIG. 3 is a block diagram showing an example architecture of a computing device that may be used in accordance with various embodiments described herein. It will be appreciated that not all devices will include all of the components of the architecture 300 and some user devices may include additional components not shown in the architecture 300. The architecture 300 may include one or more processing elements 304 for executing instructions and retrieving data stored in a storage element 302. The processing element 304 may comprise at least one processor. Any suitable processor or processors may be used. For example, the processing element 304 may comprise one or more digital signal processors (DSPs). In some examples, the processing element 304 may be effective to determine a wakeword and/or to stream audio data to a speech processing system. The storage element 302 can include one or more different types of non-transitory computer-readable memory, data storage, or computer-readable storage media devoted to different purposes within the architecture 300. For example, the storage element 302 may comprise flash memory, random-access memory, disk-based storage, etc. Different portions of the storage element 302, for example, may be used for program instructions for execution by the processing element 304, storage of images or other digital works, and/or a removable storage for transferring data to other devices, etc. In various examples, the storage element 302 may comprise instructions effective to program at least one processor to implement a privacy-preserving CV training/evaluation algorithm 388 such as the example process flow described above in reference to FIGS. 1-2F.

[0035] The storage element 302 may also store software for execution by the processing element 304. An operating system 322 may provide the user with an interface for operating the computing device and may facilitate communications and commands between applications executing on the architecture 300 and various hardware thereof. A transfer application 324 may be configured to receive images, audio, and/or video from another device (e.g., a mobile device, image capture device, and/or display device) or from an image sensor 332 and/or microphone 370 included in the architecture 300. In some examples, the transfer application 324 may also be configured to send the received voice requests to one or more voice recognition servers. Architecture 300 may store parameters and/or computer-executable instructions effective to implement the object detectors, OCR models, and/or other computer vision models, as desired.

[0036] When implemented in some user devices, the architecture 300 may also comprise a display component 306. The display component 306 may comprise one or more light-emitting diodes (LEDs) or other suitable display lamps. Also, in some examples, the display component

306 may comprise, for example, one or more devices such as cathode ray tubes (CRTs), liquid-crystal display (LCD) screens, gas plasma-based flat panel displays, LCD projectors, raster projectors, infrared projectors or other types of display devices, etc. As described herein, display component 306 may be effective to display content determined provided by a skill executed by the processing element 304 and/or by another computing device.

[0037] The architecture 300 may also include one or more input devices 308 operable to receive inputs from a user. The input devices 308 can include, for example, a push button, touch pad, touch screen, wheel, joystick, keyboard, mouse, trackball, keypad, light gun, game controller, or any other such device or element whereby a user can provide inputs to the architecture 300. These input devices 308 may be incorporated into the architecture 300 or operably coupled to the architecture 300 via wired or wireless interface. In some examples, architecture 300 may include a microphone 370 or an array of microphones for capturing sounds, such as voice requests.

[0038] When the display component 306 includes a touch-sensitive display, the input devices 308 can include a touch sensor that operates in conjunction with the display component 306 to permit users to interact with the image displayed by the display component 306 using touch inputs (e.g., with a finger or stylus). The architecture 300 may also include a power supply 314, such as a wired alternating current (AC) converter, a rechargeable battery operable to be recharged through conventional plug-in approaches, or through other approaches such as capacitive or inductive charging.

[0039] The communication interface 312 may comprise one or more wired or wireless components operable to communicate with one or more other computing devices. For example, the communication interface 312 may comprise a wireless communication module 336 configured to communicate on a network, such as a computer communication network, according to any suitable wireless protocol, such as IEEE 802.11 or another suitable wireless local area network (WLAN) protocol. A short range interface 334 may be configured to communicate using one or more short range wireless protocols such as, for example, near field communications (NFC), Bluetooth, Bluetooth LE, etc. A mobile interface 340 may be configured to communicate utilizing a cellular or other mobile protocol. A Global Positioning System (GPS) interface 338 may be in communication with one or more earth-orbiting satellites or other suitable position-determining systems to identify a position of the architecture 300. A wired communication module 342 may be configured to communicate according to the USB protocol or any other suitable protocol.

[0040] The architecture 300 may also include one or more sensors 330 such as, for example, one or more position sensors, image sensors, and/or motion sensors. An image sensor 332 is

shown in FIG. 3. An example of an image sensor 332 may be a camera configured to capture color information, image geometry information, and/or ambient light information.

[0041] FIG. 4 depicts an example process 400 for privacy preservation for computer vision model training and/or evaluation, in accordance with various aspects of the present disclosure. The actions of the process 400 may represent a series of instructions comprising computer-readable machine code executable by one or more processing units of one or more computing devices. In various examples, the computer-readable machine codes may be comprised of instructions selected from a native instruction set of and/or an operating system (or systems) of the one or more computing devices. Although the figures and discussion illustrate certain operational steps of the system in a particular order, the steps described may be performed in a different order (as well as certain steps removed or added) without departing from the intent of the disclosure.

[0042] Process 400 may begin at action 410, at which text fields in an input image may be detected (e.g., using an object recognition model (e.g., a CNN, vision transformer, R-CNN, etc.) and/or an OCR model). Any size or quantum of text fragments may be detected according to the desired implementation. For example, an object detector and/or OCR model may detect contiguous alpha-numeric characters present in the input image.

[0043] Processing may continue at action 420, at which a sub-image may be generated for each of the detected text fields. Individual sub-images may be shuffled so that they are not ordered in accordance with any specific pattern. In addition, geometric data indicating the position of the text field in the input image from which the sub-image was extracted may be stored and maintained in non-transitory computer-readable memory. As previously described, the geometric data may include 3D rotation information, two dimensional coordinate information, geometric transform data, etc.

[0044] Processing may continue at action 430, at which the generated sub-images may be sent to different annotation devices. For example, the sub-images represented the detected text fields may be sent to different annotators and/or remote computing devices so that no one annotator/remote computing device receives all the information from the input image.

[0045] At action 440, consolidated annotated image data may be generated using the location of each detected text field (e.g., as represented by text field geometric data 112), the background image, the sub-images, and the annotation for each of the detected text fields received from the annotators. Processing may continue at action 450, at which randomized alpha-numeric strings may be used to populate one or more of the text fields in the image (e.g., the background image) and/or may replace one or more alpha-numeric text strings in the image.

[0046] Processing may continue at action 460, at which attribute type annotation may be received for the text fields including the randomized alpha-numeric strings. For example, annotators may annotate the randomized text fields with attribute types describing the fields (e.g., a field including a user's name may be annotated as a "User Name" field, even though no actual user name is present in the field, only randomized text). Such attribute type annotations may be used to train an object detector to detect different attribute types corresponding to different fields for newly-input unannotated images.

[0047] **Clauses**

[0048] Embodiments of the present disclosure can be described in view of the following clauses:

Clause 1. A computer-implemented method comprising:

- detecting, by an optical character recognition (OCR) component, at least a first field of contiguous text in a first image and a second field of contiguous text in the first image;

- determining a first alpha-numeric text string in the first field;

- determining a second alpha-numeric text string in the second field, wherein at least one of the first alpha-numeric text string and the second alpha-numeric text string comprises personally-identifiable information (PII);

- generating a first sub-image comprising the first alpha-numeric text string;

- generating a second sub-image comprising the second alpha-numeric text string;

- receiving, from a first annotator, first label data for the first sub-image;

- receiving, from a second annotator, second label data for the second sub-image; and

- training a first computer vision model based at least in part on:

- generating a first prediction by the first computer vision model for the first sub-image;

- comparing the first prediction to the first label data;

- generating a second prediction by the first computer vision model for the second sub-image; and

- comparing the second prediction to the second label data.

Clause 2. The computer-implemented method of clause 1, further comprising:

- generating second image data, wherein the second image data represents a background of the first image with the first alpha-numeric text string removed from the first field and the second alpha-numeric text string removed from the second field; and

- receiving a first bounding box annotation from the first annotator, the first bounding box annotation representing a text field present in the first image undetected by the OCR component.

Clause 3. The computer-implemented method of clause 1, further comprising:

generating second image data, wherein the second image data represents a background of the first image with the first alpha-numeric text string removed from the first field and the second alpha-numeric text string removed from the second field; and

populating the first field in the second image data with a first randomized alpha-numeric text string and the second field in the second image data with a second randomized alpha-numeric text string.

Clause 4. The computer-implemented method of clause 3, further comprising:

receiving a first bounding box annotation from the first annotator for the second image data, the first bounding box annotation defining an attribute type for the first field.

Clause 5. A computer-implemented method comprising:

detecting a first field of text in a first image and a second field of text in the first image;

determining a first alpha-numeric text string in the first field;

determining a second alpha-numeric text string in the second field;

generating a first sub-image comprising the first alpha-numeric text string;

generating a second sub-image comprising the second alpha-numeric text string;

sending the first sub-image to a first computing device for annotation; and

sending the second sub-image to a second computing device for annotation.

Clause 6. The computer-implemented method of clause 5, further comprising:

generating second image data representing a background of the first image with text removed;

determining first bounding box data representing a location of the first field in the second image data; and

determining second bounding box data representing a location of the second field in the second image data.

Clause 7. The computer-implemented method of clause 6, further comprising receiving, from the first computing device, a first annotation representing an undetected text field in the second image data.

Clause 8. The computer-implemented method of clause 6, further comprising:

generating a third alpha-numeric text string in the first field in the second image data, wherein the third alpha-numeric text string comprises pseudo-random characters; and

generating a fourth alpha-numeric text string in the second field in the second image data, wherein the third alpha-numeric text string comprises pseudo-random characters.

Clause 9. The computer-implemented method of clause 8, further comprising:

saving the second image data with the third alpha-numeric text string in the first field and the fourth alpha-numeric text string in the second field as third image data, wherein the third image data represents formatting of the first image with different text strings replacing text in detected fields.

Clause 10. The computer-implemented method of clause 9, further comprising:

sending the third image data to the first computing device;

receiving, from the first computing device, a first bounding box annotation for the first field of the third image data, the first bounding box annotation labeled with a first attribute type of the first field; and

receiving, from the first computing device, a second bounding box annotation for the second field of the third image data, the second bounding box annotation labeled with a second attribute type of the second field.

Clause 11. The computer-implemented method of clause 5, wherein the first alpha-numeric text string in the first sub-image comprises less than a total amount of text present in the first field of text in the first image.

Clause 12. The computer-implemented method of clause 5, further comprising:

determining geometric data representing a location of the first field of text in the first image; and

generating second image data representing a background of the first image with text removed, wherein the location of the first field of text in the second image data is determined using the geometric data.

Clause 13. A system comprising:

at least one processor; and

non-transitory computer-readable memory storing instructions that, when executed by the at least one processor, are effective to cause the at least one processor to:

- detect a first field of text in a first image and a second field of text in the first image;
- determine a first alpha-numeric text string in the first field;
- determine a second alpha-numeric text string in the second field;
- generate a first sub-image comprising the first alpha-numeric text string;
- generate a second sub-image comprising the second alpha-numeric text string;
- send the first sub-image to a first computing device for annotation; and
- send the second sub-image to a second computing device for annotation.

Clause 14. The system of clause 13, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

- generate second image data representing a background of the first image with text removed;
- determine first bounding box data representing a location of the first field in the second image data; and
- determine second bounding box data representing a location of the second field in the second image data.

Clause 15. The system of clause 14, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to receive, from the first computing device, a first annotation representing an undetected text field in the second image data.

Clause 16. The system of clause 14, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

- generate a third alpha-numeric text string in the first field in the second image data, wherein the third alpha-numeric text string comprises pseudo-random characters; and
- generate a fourth alpha-numeric text string in the second field in the second image data, wherein the third alpha-numeric text string comprises pseudo-random characters.

Clause 17. The system of clause 16, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

save the second image data with the third alpha-numeric text string in the first field and the fourth alpha-numeric text string in the second field as third image data in the non-transitory computer-readable memory, wherein the third image data represents formatting of the first image with different text strings replacing text in detected fields.

Clause 18. The system of clause 17, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

- send the third image data to the first computing device;

- receive, from the first computing device, a first bounding box annotation for the first field of the third image data, the first bounding box annotation labeled with a first attribute type of the first field; and

- receive, from the first computing device, a second bounding box annotation for the second field of the third image data, the second bounding box annotation labeled with a second attribute type of the second field.

Clause 19. The system of clause 13, wherein the first alpha-numeric text string in the first sub-image comprises less than a total amount of text present in the first field of text in the first image.

Clause 20. The system of clause 13, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

- determine geometric data representing a location of the first field of text in the first image; and

- generate second image data representing a background of the first image with text removed, wherein the location of the first field of text in the second image data is determined using the geometric data.

[0049] Although various systems described herein may be embodied in software or code executed by general purpose hardware as discussed above, as an alternate the same may also be embodied in dedicated hardware or a combination of software/general purpose hardware and dedicated hardware. If embodied in dedicated hardware, each can be implemented as a circuit or state machine that employs any one of or a combination of a number of technologies. These technologies may include, but are not limited to, discrete logic circuits having logic gates for

implementing various logic functions upon an application of one or more data signals, application specific integrated circuits having appropriate logic gates, or other components, etc. Such technologies are generally well known by those of ordinary skill in the art and consequently, are not described in detail herein.

[0050] The flowcharts and methods described herein show the functionality and operation of various implementations. If embodied in software, each block or step may represent a module, segment, or portion of code that comprises program instructions to implement the specified logical function(s). The program instructions may be embodied in the form of source code that comprises human-readable statements written in a programming language or machine code that comprises numerical instructions recognizable by a suitable execution system such as a processing component in a computer system. If embodied in hardware, each block may represent a circuit or a number of interconnected circuits to implement the specified logical function(s).

[0051] Although the flowcharts and methods described herein may describe a specific order of execution, it is understood that the order of execution may differ from that which is described. For example, the order of execution of two or more blocks or steps may be scrambled relative to the order described. Also, two or more blocks or steps may be executed concurrently or with partial concurrence. Further, in some embodiments, one or more of the blocks or steps may be skipped or omitted. It is understood that all such variations are within the scope of the present disclosure.

[0052] Also, any logic or application described herein that comprises software or code can be embodied in any non-transitory computer-readable medium or memory for use by or in connection with an instruction execution system such as a processing component in a computer system. In this sense, the logic may comprise, for example, statements including instructions and declarations that can be fetched from the computer-readable medium and executed by the instruction execution system. In the context of the present disclosure, a "computer-readable medium" can be any medium that can contain, store, or maintain the logic or application described herein for use by or in connection with the instruction execution system. The computer-readable medium can comprise any one of many physical media such as magnetic, optical, or semiconductor media. More specific examples of a suitable computer-readable media include, but are not limited to, magnetic tapes, magnetic floppy diskettes, magnetic hard drives, memory cards, solid-state drives, USB flash drives, or optical discs. Also, the computer-readable medium may be a random access memory (RAM) including, for example, static random access memory (SRAM) and dynamic random access memory (DRAM), or magnetic random access memory (MRAM). In addition, the computer-readable medium may be a read-only memory (ROM), a programmable read-only memory (PROM), an erasable programmable read-only

memory (EPROM), an electrically erasable programmable read-only memory (EEPROM), or other type of memory device.

[0053] It should be emphasized that the above-described embodiments of the present disclosure are merely possible examples of implementations set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described example(s) without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and protected by the following claims.

CLAIMS

WHAT IS CLAIMED IS:

1. A computer-implemented method comprising:
 - detecting a first field of text in a first image and a second field of text in the first image;
 - determining a first alpha-numeric text string in the first field;
 - determining a second alpha-numeric text string in the second field;
 - generating a first sub-image comprising the first alpha-numeric text string;
 - generating a second sub-image comprising the second alpha-numeric text string;
 - sending the first sub-image to a first computing device for annotation; and
 - sending the second sub-image to a second computing device for annotation.
2. The computer-implemented method of claim 1, further comprising:
 - generating second image data representing a background of the first image with text removed;
 - determining first bounding box data representing a location of the first field in the second image data; and
 - determining second bounding box data representing a location of the second field in the second image data.
3. The computer-implemented method of claim 2, further comprising receiving, from the first computing device, a first annotation representing an undetected text field in the second image data.
4. The computer-implemented method of claim 2, further comprising:
 - generating a third alpha-numeric text string in the first field in the second image data, wherein the third alpha-numeric text string comprises pseudo-random characters; and
 - generating a fourth alpha-numeric text string in the second field in the second image data, wherein the fourth alpha-numeric text string comprises pseudo-random characters.
5. The computer-implemented method of claim 4, further comprising:
 - saving the second image data with the third alpha-numeric text string in the first field and the fourth alpha-numeric text string in the second field as third image data, wherein the third image data represents formatting of the first image with different text strings replacing text in detected fields.

6. The computer-implemented method of claim 5, further comprising:
 - sending the third image data to the first computing device;
 - receiving, from the first computing device, a first bounding box annotation for the first field of the third image data, the first bounding box annotation labeled with a first attribute type of the first field; and
 - receiving, from the first computing device, a second bounding box annotation for the second field of the third image data, the second bounding box annotation labeled with a second attribute type of the second field.
7. The computer-implemented method of claim 1 or 2, wherein the first alpha-numeric text string in the first sub-image comprises less than a total amount of text present in the first field of text in the first image.
8. The computer-implemented method of claim 1, further comprising:
 - determining geometric data representing a location of the first field of text in the first image; and
 - generating second image data representing a background of the first image with text removed, wherein the location of the first field of text in the second image data is determined using the geometric data.
9. 13. A system comprising:
 - at least one processor; and
 - non-transitory computer-readable memory storing instructions that, when executed by the at least one processor, are effective to cause the at least one processor to:
 - detect a first field of text in a first image and a second field of text in the first image;
 - determine a first alpha-numeric text string in the first field;
 - determine a second alpha-numeric text string in the second field;
 - generate a first sub-image comprising the first alpha-numeric text string;
 - generate a second sub-image comprising the second alpha-numeric text string;
 - send the first sub-image to a first computing device for annotation; and
 - send the second sub-image to a second computing device for annotation.

10. The system of claim 9, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

generate second image data representing a background of the first image with text removed;

determine first bounding box data representing a location of the first field in the second image data; and

determine second bounding box data representing a location of the second field in the second image data.

11. The system of claim 10, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to receive, from the first computing device, a first annotation representing an undetected text field in the second image data.

12. The system of claim 10, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

generate a third alpha-numeric text string in the first field in the second image data, wherein the third alpha-numeric text string comprises pseudo-random characters; and

generate a fourth alpha-numeric text string in the second field in the second image data, wherein the third alpha-numeric text string comprises pseudo-random characters.

13. The system of claim 12, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

save the second image data with the third alpha-numeric text string in the first field and the fourth alpha-numeric text string in the second field as third image data in the non-transitory computer-readable memory, wherein the third image data represents formatting of the first image with different text strings replacing text in detected fields.

14. The system of claim 13, the non-transitory computer-readable memory storing further instructions that, when executed by the at least one processor, are further effective to cause the at least one processor to:

send the third image data to the first computing device;

receive, from the first computing device, a first bounding box annotation for the first field of the third image data, the first bounding box annotation labeled with a first attribute type of the first field; and

receive, from the first computing device, a second bounding box annotation for the second field of the third image data, the second bounding box annotation labeled with a second attribute type of the second field.

15. The system of claim 9 or 10, wherein the first alpha-numeric text string in the first sub-image comprises less than a total amount of text present in the first field of text in the first image.

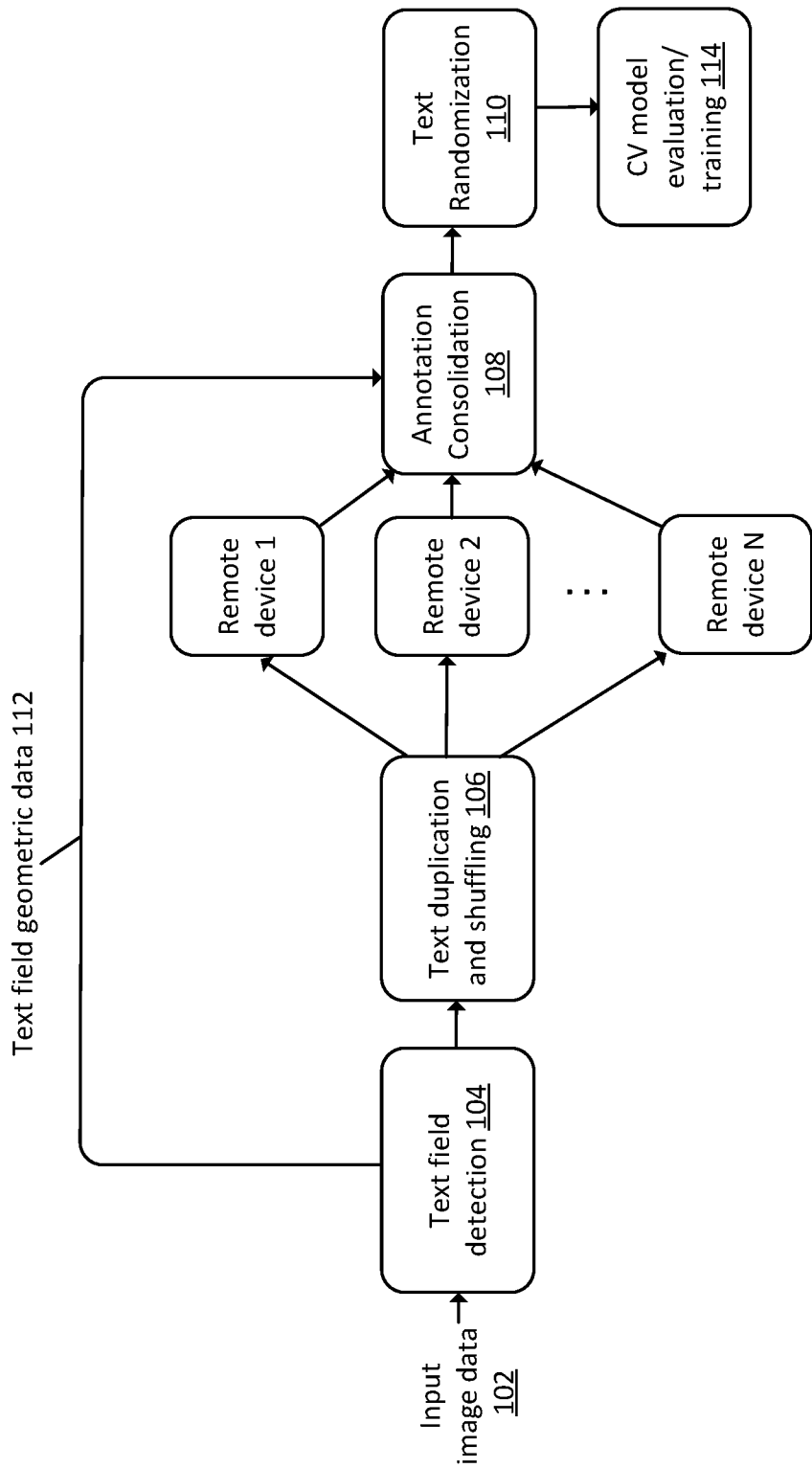


FIG. 1

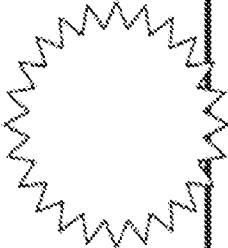
INSURCO HEALTHSHIELD	 HEALTHSHIELD Tier 1 
John Q. Member Identification Number JQP123X45678	
Effective Date Contract Code Rx Bin PCN Rx Group Plan	5000 Ind-10000 Fam 10000 Ind -20000 Fam 70% 50%
Select Rx List	
Dental Program: Prime	Pathway X PPO  Rx

FIG. 2A

INSURCO

HEALTHSHIELD

John Q

Member

Identification Number

JQP123X45678

Effective Date

Contract Code

Rx Bin

PCN

Rx Group

Plan

Select Rx List

Dental Program: Prime

HEALTHSHIELD

Tier 1

Deductible Out

Co-Insurance In

Co-Insurance Out

Ded In Network

5000 Ind - 10000 Fam

10000 Ind - 20000 Fam

70%

50%

Pathway X PPO

PPOB

Rx

FIG. 2B

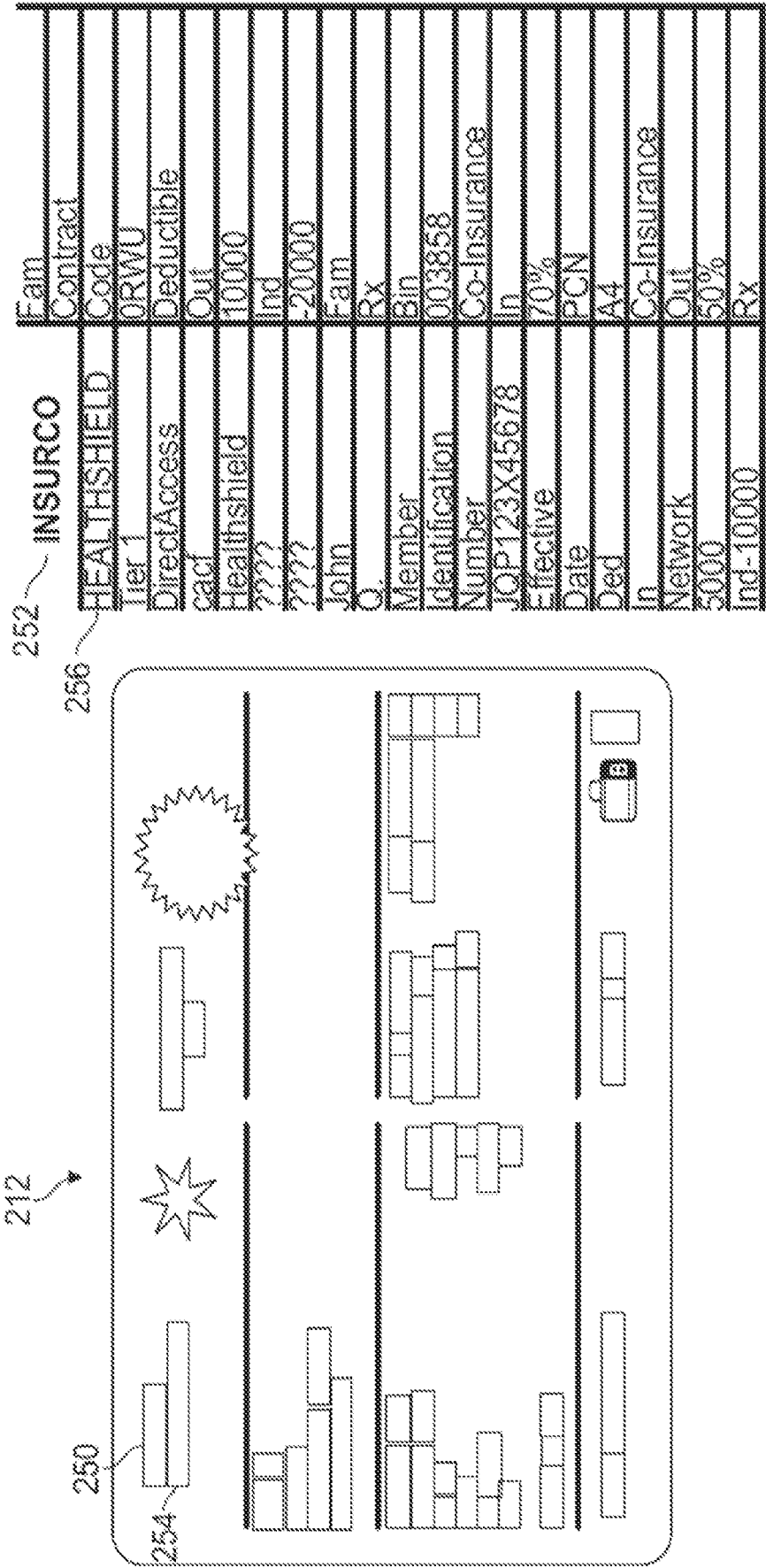


FIG. 2C

If the text in the left image is the same as the text in the right image label 'True' otherwise label 'False'

☐ True

☐ False

FIG. 2D

Type the text that you see in the display on the left in the field below:

FIG. 2E

JQP123X45678 JQP123X45678

WLHA WLHA

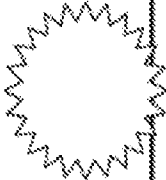
WLHA

INSURCO

HEALTHSHIELD



HEALTHSHIELD
Tier 1



John Q.
Member
Identification Number
[SS559C03936]

Effective Date
Contract Code
Rx Bin
PCN
Rx Group
Plan

0RWU
-877587
A4
[PRKL]
040

Select Rx List

Dental Program: Prime

Ded In Network
Deductible Out
Co-Insurance In
Co-Insurance Out

5000 Ind-10000 Fam
10000 Ind -20000 Fam
70%
50%

Pathway X PPO

 Rx

Draw a polygon label around the following fields:

Member ID: _____

Bin_Number: - - - - -

Group_ID: - - - - -

*
*
*

FIG. 2F

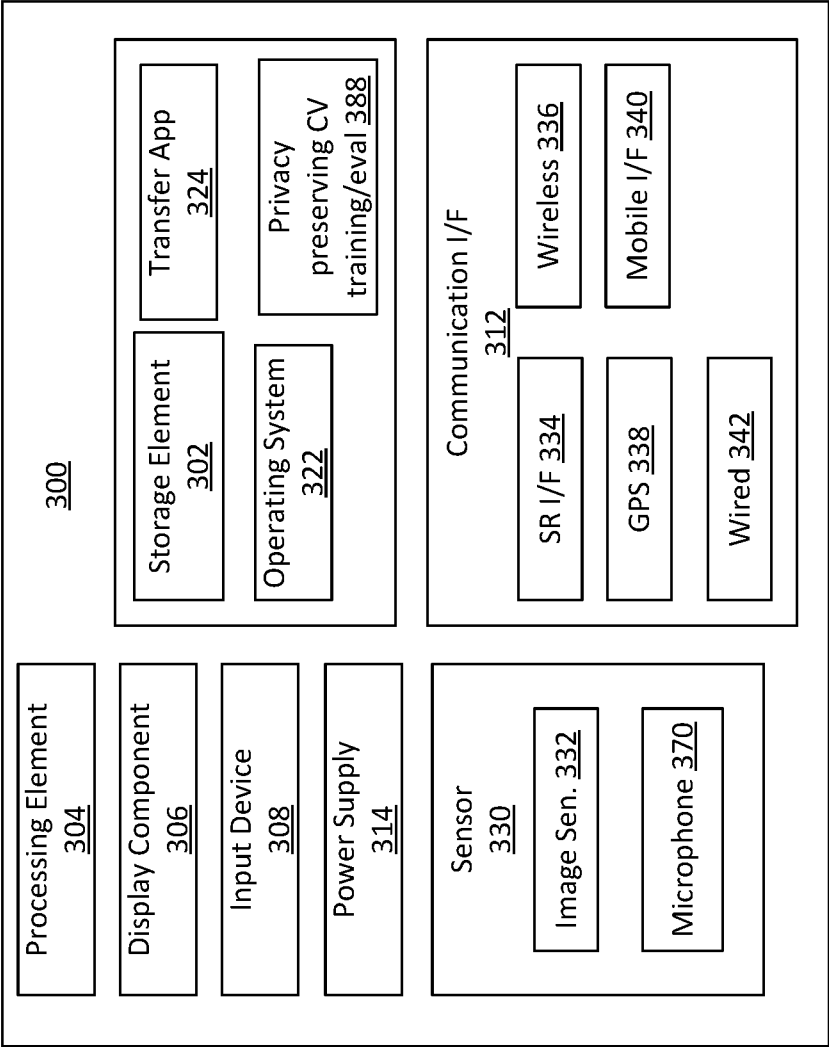
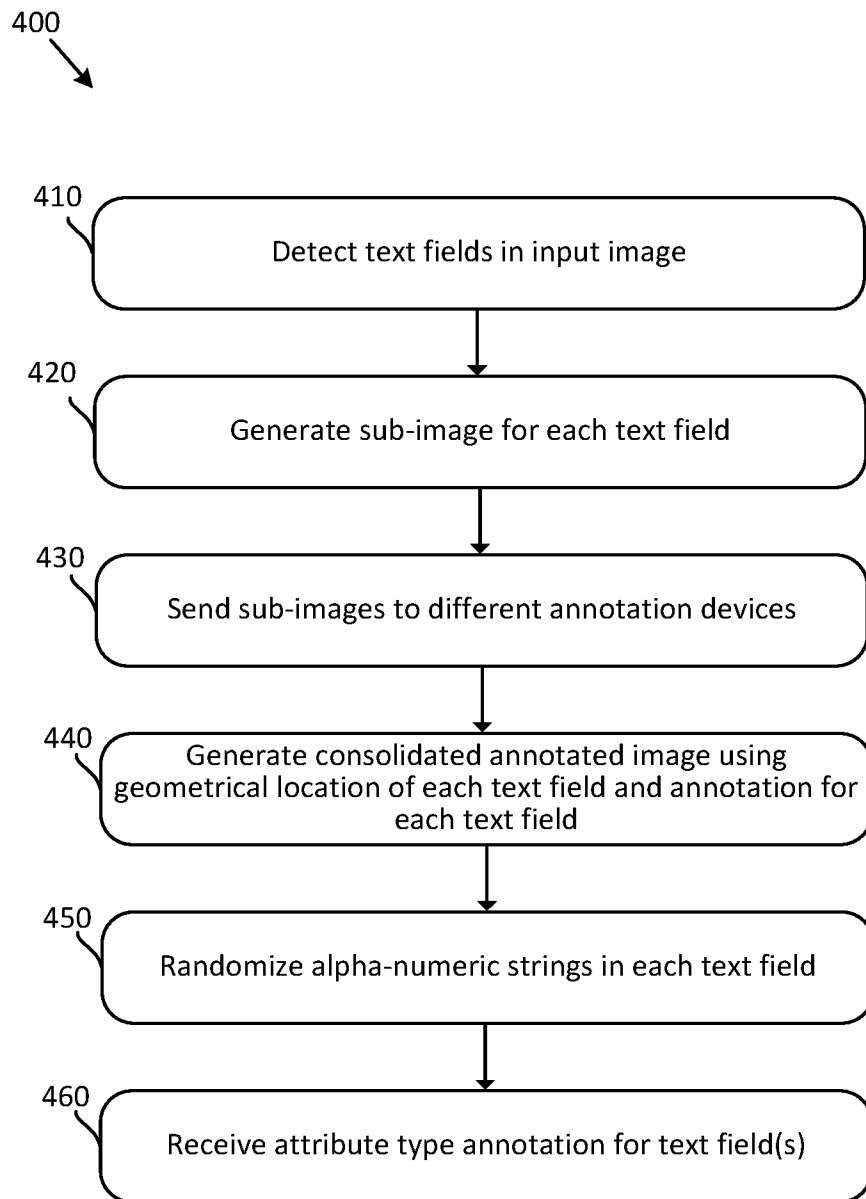


FIG. 3

8 / 8

**FIG. 4**

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/033282

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06V30/412 G06V30/19 G06V30/414 G06F21/62
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06V G06F

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2021/357512 A1 (BUSILA ELENA [CA] ET AL) 18 November 2021 (2021-11-18) paragraph [0006] paragraph [0031] - paragraph [0034] paragraph [0043] - paragraph [0048] ----- - / - -	1 - 15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

13 September 2024

Date of mailing of the international search report

24/09/2024

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Rajadell Rojas, Olga

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/033282

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DE CAPITANI DI VIMERCATI SABRINA ET AL: "Artifact: Scalable Distributed Data Anonymization", 2021 IEEE INTERNATIONAL CONFERENCE ON PERVASIVE COMPUTING AND COMMUNICATIONS WORKSHOPS AND OTHER AFFILIATED EVENTS (PERCOM WORKSHOPS), IEEE, 22 March 2021 (2021-03-22), pages 450-451, XP033917117, DOI: 10.1109/PERCOMWORKSHOPS51409.2021.9431059 [retrieved on 2021-05-13] Section I.Introduction; page 1	1-15
Y	US 2021/397737 A1 (DE BERKER ARCHY OTTO [CA] ET AL) 23 December 2021 (2021-12-23) paragraph [0005] paragraph [0042] paragraph [0048] - paragraph [0053] paragraph [0058] - paragraph [0064]	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2024/033282

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2021357512 A1	18-11-2021	AU 2019366169 A1	10-06-2021
		CA 3117374 A1	30-04-2020
		CN 113228019 A	06-08-2021
		EP 3871125 A1	01-09-2021
		JP 7418423 B2	19-01-2024
		JP 2022505908 A	14-01-2022
		KR 20210095875 A	03-08-2021
		US 2021357512 A1	18-11-2021
US 2021397737 A1	23-12-2021	WO 2020082187 A1	30-04-2020
		AU 2019374742 A1	03-06-2021
		CN 113272806 A	17-08-2021
		EP 3877885 A1	15-09-2021
		JP 7353366 B2	29-09-2023
		JP 2022506866 A	17-01-2022
		KR 20210102240 A	19-08-2021
		US 2021397737 A1	23-12-2021
		WO 2020093165 A1	14-05-2020