

METHOD AND SYSTEM FOR DETERMINING AUTHENTICITY OF AN OFFICIAL DOCUMENT

TECHNICAL FIELD

[0001] The present specification relates broadly, but not exclusively, to methods, systems and devices for determining authenticity of an official document.

BACKGROUND

[0002] Know Your Customer (KYC) is a form of due diligence that a business does in order to verify the authenticity, identity, risks, etc. of its customers/clients either before or during the time that it starts doing business with them. Electronic Know Your Customer (eKYC), or digital KYC, is a paperless process that is performed by the business, usually through use of a computer, and may sometimes be initiated by a customer. In the eKYC process, there are typical steps including scanning an identity (ID) card and detecting a live face. During the process of scanning an ID card, it is usually necessary to check the authenticity of the ID card.

[0003] One fake ID card detection method involves using similarity scores to detect security features, such as landmarks or holograms, to verify the authenticity of the ID card. While this method works well for a certain type of ID card, e.g. one where the security features are concentrated together, it has been noted that the designs of other types of ID cards may have security features that vary significantly in the number, size, distribution, etc.

[0004] A need therefore exists to provide alternative or improved methods and devices for determining authenticity of an official document.

SUMMARY

[0005] An embodiment provides a method of determining authenticity of an official document. The method includes extracting image data of a plurality of security features of the official document from an image of the official document; applying a trained neural network to the extracted image data to determine a sharpness of each of the security features on the image of the official document; and determining whether the official document is authentic based on a sharpness classification of the security features. Training the neural network includes training a multi-label classification model to classify a plurality of regions of interest, wherein the regions of interests are spatially distributed on a labeled training image, and wherein each label corresponds to a respective one of the regions of interest; applying a spatial attention module to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest; and optimizing parameters of the neural network based on optimizing a loss of the multi-label classification model over a plurality of labeled training images. The labeled training images include labeled images of official documents, and the regions of interest include selected cropped regions containing the security features on said official documents.

[0006] Another embodiment provides a system for determining authenticity of an official document. The system includes a processor and a computer-readable memory coupled to the processor and having instructions stored thereon that are executable by the processor to extract image data of a plurality of security features of the official document from an image of the official document; apply a trained neural network to the extracted image data to determine a sharpness of each of the security features on the image of the official document; and determine whether the official document is authentic based on a sharpness classification of the security features. The trained neural network includes a multi-label classification model trained to classify a plurality of regions of interest, wherein the regions of interests are spatially distributed on a labeled training image, and wherein each label corresponds to a respective one of the regions of interest; a spatial attention module applied to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest; and neural network parameters optimized based on an optimization of a loss of the multi-label classification over a plurality of labeled training images. The labeled training images include labeled images of official documents, and the regions of interest include selected cropped regions containing the security features on said official documents.

[0007] Another embodiment provides an apparatus comprising an input device and a processing device. The input device can receive an image of the official

document. The processing device can for extract image data of a plurality of security features of the official document from the image of the official document and feed the extracted image data to a trained neural network to determine: a sharpness of each of the security features on the image of the official document and an authenticity of the official document based on a sharpness classification of the security features. The trained neural network includes a multi-label classification model trained to classify a plurality of regions of interest, wherein the regions of interests are spatially distributed on a labeled training image, and wherein each label corresponds to a respective one of the regions of interest; a spatial attention module applied to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest; and neural network parameters optimized based on an optimization of a loss of the multi-label classification over a plurality of labeled training images. The labeled training images include labeled images of official documents, and the regions of interest include selected cropped regions containing the security features on said official documents.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Embodiments will be better understood and readily apparent to one of ordinary skill in the art from the following written description, by way of example only, and in conjunction with the drawings, in which:

[0009] Figure 1 shows a flowchart illustrating a method of determining authenticity of an official document according to an embodiment.

[0010] Figure 2 shows a flowchart illustrating a detailed example implementation of the method of Figure 1.

[0011] Figure 3 shows a schematic diagram of an apparatus suitable for implementing the method of Figure 1.

[0012] Figure 4 shows a schematic diagram illustrating a computer system suitable for implementing the method of Figure 1.

[0013] Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been depicted to scale. For example, the dimensions of some of the elements in the illustrations, block diagrams or

flowcharts may be exaggerated in respect to other elements to help to improve understanding of the present embodiments.

DETAILED DESCRIPTION

[0014] The present specification provides methods, systems and devices to determine whether an official document, such as an identity card, is authentic based on an image of the official document. Briefly, through machine learning, an attention-based multi-label classification model, which pays more attention to regions of interest (ROIs) on the image with security features during the training process, is developed and then deployed in determining the authenticity of the official document. A spatial attention module is applied to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest. The attention-based approach in the present method can neglect the effects of other irrelevant information on the image so as to improve accuracy. For example, the present method has been validated on a new ID card design and achieved 95% accuracy on very challenging datasets. In other words, the method not only can distinguish between genuine and fake official documents, but also accommodate different designs or design iterations of the official documents.

[0015] Embodiments will be described, by way of example only, with reference to the drawings. Like reference numerals and characters in the drawings refer to like elements or equivalents.

[0016] Some portions of the description which follows are explicitly or implicitly presented in terms of algorithms and functional or symbolic representations of operations on data within a computer memory. These algorithmic descriptions and functional or symbolic representations are the means used by those skilled in the data processing arts to convey most effectively the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transferred, combined, compared, and otherwise manipulated.

[0017] Unless specifically stated otherwise, and as apparent from the following, it will be appreciated that throughout the present specification, discussions utilizing terms

such as “calculating”, “determining”, “training”, “applying”, “extracting”, “generating”, “optimizing”, “outputting”, or the like, refer to the action and processes of a computer system, or similar electronic device, that manipulates and transforms data represented as physical quantities within the computer system into other data similarly represented as physical quantities within the computer system or other information storage, transmission or display devices.

[0018] The present specification also discloses apparatus for performing the operations of the methods. Such apparatus may be specially constructed for the required purposes, or may comprise a computer or other device selectively activated or reconfigured by a computer program stored in the computer. The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various machines may be used with programs in accordance with the teachings herein. Alternatively, the construction of more specialized apparatus to perform the required method steps may be appropriate. The structure of a conventional computer will appear from the description below.

[0019] In addition, the present specification also implicitly discloses a computer program, in that it would be apparent to the person skilled in the art that the individual steps of the method described herein may be put into effect by computer code. The computer program is not intended to be limited to any particular programming language and implementation thereof. It will be appreciated that a variety of programming languages and coding thereof may be used to implement the teachings contained herein. Moreover, the computer program is not intended to be limited to any particular control flow. There are many other variants of the computer program, which can use different control flows without departing from the scope of the specification.

[0020] Furthermore, one or more of the steps of the computer program may be performed in parallel rather than sequentially. Such a computer program may be stored on any computer readable medium. The computer readable medium may include storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a computer. The computer readable medium may also include a hard-wired medium such as exemplified in the Internet system, or wireless medium such as exemplified in the GSM, GPRS, 3G, 4G or 5G mobile telephone systems, as well as other wireless systems such as Bluetooth, ZigBee, Wi-Fi. The computer program when loaded and executed on such a computer effectively results in an apparatus that implements the steps of the preferred method.

[0021] The present specification may also be implemented as hardware elements. More particularly, in the hardware sense, an element is a functional hardware unit designed for use with other components or elements. For example, an element may be implemented using discrete electronic components, or it can form a portion of an entire electronic circuit such as an Application Specific Integrated Circuit (ASIC) or Field Programmable Gate Array (FPGA). Numerous other possibilities exist. Those skilled in the art will appreciate that the system can also be implemented as a combination of hardware and software elements.

[0022] According to various embodiments, a “circuit” may be understood as any kind of a logic implementing entity, which may be special purpose circuitry or a processor executing software stored in a memory, firmware, or any combination thereof. Thus, in an embodiment, a “circuit” may be a hard-wired logic circuit or a programmable logic circuit such as a programmable processor, e.g. a microprocessor (e.g. a Complex Instruction Set Computer (CISC) processor or a Reduced Instruction Set Computer (RISC) processor). A “circuit” may also be a processor executing software, e.g. any kind of computer program, e.g. a computer program using a virtual machine code such as e.g. Java. Any other kind of implementation of the respective functions which may be described in more detail herein may also be understood as a “circuit” in accordance with an alternative embodiment.

[0023] Figure 1 shows a flowchart 100 illustrating a method for determining authenticity of an official document, or an officially-issued document, according to an embodiment.

[0024] At step 102, image data of a plurality of security features of the official document is extracted from an image of the official document. For example, the extraction may be performed by identifying the positions of the security features and subtracting or cropping portions of the image without the security features.

[0025] The official document may be, for example, an identity card as discussed in further details below with reference to Figure 2. Other non-limiting examples of a suitable official document include a passport, a driving license, a social security card, a certificate (birth certificate, marriage certificate, education certificate, etc.), a cheque, a bank note, etc. Such an official document typically has pre-defined fields and features including the security features (e.g. holograms, patterns,

shapes, etc.). Once the position of the official document in the image is known, the positions of the security features can be determined.

[0026] At step 104, a trained neural network is applied to the extracted image data to determine a sharpness of each of the security features on the image of the official document.

[0027] In the training stage, a multi-label classification model is trained to classify a plurality of regions of interest. The regions of interests are spatially distributed on a labeled training image, and each label of the classification model corresponds to a respective one of the regions of interest. In addition, a spatial attention module is applied to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest. Then, parameters of the neural network are optimized based on optimizing a loss of the multi-label classification model over a plurality of labeled training images.

[0028] Moreover, in the training stage, the labeled training images include labeled images of official documents, the regions of interest include regions containing the security features on such official documents, and the at least one attribute includes the sharpness of the security features on the labeled images of official documents.

[0029] Accordingly, an output from step 104 is a sharpness classification of each of the security features on the image of the official document.

[0030] At step 106, whether the official document is authentic is determined based on a sharpness classification of the security features. For example, if the official document is an identity card and, for security reasons, a threshold may be set such that the identity card is determined to be authentic only if each of the security features on the image of identity card is classified as sharp. In other words, if any of the security features is absent or out of position, or if there is a hint of manipulation or tampering to any one of the security features, the identity card is not considered to be authentic.

[0031] Figure 2 shows a flowchart 200 illustrating a detailed implementation of the method of Figure 1. In this embodiment, the official document is an identity card 202 (hereinafter interchangeably referred to as ID card 202), but it will be appreciated by a person skilled in the art that the method can be similarly applied to other types of official documents.

[0032] At step 204, an ID card image I1 which may be a photo taken by the user by using, for example, a camera-enabled cell phone, is provided. Typically, the user only needs to provide one ID card image I1 taken at a single angle. If the ID card 202 is placed on a table, for example, the ID card image I1 may include both the ID card 202 and the table surface as background. Alternatively, the ID card image I1 may be a scan obtained from a scanner in other embodiments.

[0033] To obtain the security features, one or more image subtraction or cropping steps are performed on the card image I1. In one embodiment, the four corners C1, C2, C3, C4 of the ID card 202 are detected at step 206, and an image area I2 corresponding to the ID card 202 is extracted by removing the background at step 208. Next, the extracted imaged area I2 is further cropped to retain a plurality of regions of interest R1, R2, R3, R4 containing the plurality of security features at step 210. For example, for an 8-bit image, cropping can be carried out by setting each pixel in the regions with the security features to a value between 1 and 255, and setting the pixels in unrelated/irrelevant regions to a value 0. In this embodiment, the regions of interest R1, R2, R3, R4 with security features correspond to the 'See through windows', 'Hologram', 'Left triangles' and 'Right waves' features, respectively, whose positions on a front side of the ID card 202 are pre-defined.

[0034] In alternate embodiments, it will be appreciated that a different number of regions may be used, and that images of both the front and back sides of the ID card may be used to extract the regions of interest.

[0035] Next, at step 212, the extracted regions of interest R1, R2, R3, R4 are fed to a trained neural network, for example, a trained convolutional neural network 214, to determine their sharpness based on a multi-label classification. The trained neural network 214 includes a multi-label classification model trained to classify a plurality of regions of interest, with the regions of interests being spatially distributed on a labeled training image and each label of the multi-label classification model corresponding to a respective one of the regions of interest. A spatial attention module is applied to at least one layer of the neural network to learn a position and at least one attribute (e.g. sharpness) of each of the regions of interest. It is noted that the spatial positions of learned features are not consistent in a cropped document. Thus, an attention module, comprising channel attention and spatial attention, is adopted to enable the neural network 214 to focus more on relevant elements of the input than on irrelevant parts. The spatial attention focuses on 'where' is an informative part, which is complementary to the channel

attention. For example, the spatial attention module is applied to the last fully-connected layer of the neural network 214, such that features vectors from the last fully-connected layer are extracted as the representation vectors of the regions of interest, then both max pooling and average pooling are adopted to output a shared network. The neural network parameters are optimized based on an optimization of a loss of the multi-label classification over a plurality of labeled training images.

[0036] An output of step 212 is a classification value for each of the extracted regions. For example, a classification value of 1 (i.e. true) is assigned if the respective region is considered sharp. In one embodiment, the ID card 202 may be determined to be authentic if each of the regions R1, R2, R4, R4 containing the security features is classified as sharp. This can help to ensure compliance with a strict protocol regarding verification of ID cards. In other embodiments, e.g. where a lower threshold may be applied, the regions may be ranked or weighted such that a document may be verified as authentic if selected regions (e.g. more important ones) are classified as sharp. One way to implement that is to calculate a weighted sum of the classification values and base the classification on the value of the weighted sum.

[0037] In one embodiment, at least some of the steps as described above are performed using elements of the Convolutional Architecture for Fast Feature Embedding (CAFFE) framework. It will be appreciated that other deep learning frameworks may be used in alternate embodiments.

[0038] Figure 3 shows a schematic diagram of an apparatus 300 suitable for implementing the method of Figure 1. The apparatus 300 includes a receiving/input device 302 and a processing device 304. The receiving device 302 can receive an image of the official document. The processing device 304 can extract image data of a plurality of security features of the official document from the image of the official document and feed the extracted image data to a trained neural network 306. The trained neural network 306 includes a multi-label classification model 308 trained to classify a plurality of regions of interest, with the regions of interests being spatially distributed on a labeled training image, and each label of the classification model corresponding to a respective one of the regions of interest. A spatial attention module 310 is applied to at least one layer of the neural network 306 to learn a position and at least one attribute of each of the regions of interest.

Neural network parameters 312 are optimized based on an optimization of a loss of the multi-label classification over a plurality of labeled training images. The trained neural network 306 can determine a sharpness of each of the security features on the image of the official document, and the authenticity of the official document can then be verified based on a sharpness classification of the security features. For example, if the official document is an ID card, a positive verification is provided by the processing device 304 only if each of the security features on the image of the official document is classified as sharp.

[0039] As described, the present methods, systems and devices make use of a spatial attention module to allow the neural network to focus on regions of interest which are spatially distributed on an image. Thus, speed and accuracy of the classification can be improved. In addition, the present approach can be quickly adapted to verify authenticity of a different design of official document which has a different spatial distribution of the security features. The multi-label classification model can also accommodate changes to the number of regions of interest. Furthermore, rigorousness and security levels of the final authenticity determination can be provided through the number of true/positive classification values from among the regions of interest. In other words, overall performance is improved compared to existing methods.

[0040] Figure 4 depicts an exemplary computing device 400, hereinafter interchangeably referred to as a computer system 400, where one or more such computing devices 400 may be used for apparatus 300. The following description of the computing device 400 is provided by way of example only and is not intended to be limiting.

[0041] As shown in Figure 4, the example computing device 400 includes a processor 404 for executing software routines. Although a single processor is shown for the sake of clarity, the computing device 400 may also include a multi-processor system. The processor 404 is connected to a communication infrastructure 406 for communication with other components of the computing device 400. The communication infrastructure 406 may include, for example, a communications bus, cross-bar, or network.

[0042] The computing device 400 further includes a main memory 408, such as a random access memory (RAM), and a secondary memory 410. The secondary memory 410 may include, for example, a hard disk drive 412 and/or a removable storage drive 414, which may include a floppy disk drive, a magnetic tape drive, an

optical disk drive, or the like. The removable storage drive 414 reads from and/or writes to a removable storage unit 418 in a well-known manner. The removable storage unit 418 may include a floppy disk, magnetic tape, optical disk, or the like, which is read by and written to by removable storage drive 414. As will be appreciated by persons skilled in the relevant art(s), the removable storage unit 418 includes a computer readable storage medium having stored therein computer executable program code instructions and/or data.

[0043] In an alternative implementation, the secondary memory 410 may additionally or alternatively include other similar means for allowing computer programs or other instructions to be loaded into the computing device 400. Such means can include, for example, a removable storage unit 422 and an interface 420. Examples of a removable storage unit 422 and interface 420 include a program cartridge and cartridge interface (such as that found in video game console devices), a removable memory chip (such as an EPROM or PROM) and associated socket, and other removable storage units 422 and interfaces 420 which allow software and data to be transferred from the removable storage unit 422 to the computer system 400.

[0044] The computing device 400 also includes at least one communication interface 424. The communication interface 424 allows software and data to be transferred between computing device 400 and external devices via a communication path 426. In various embodiments, the communication interface 424 permits data to be transferred between the computing device 400 and a data communication network, such as a public data or private data communication network. The communication interface 424 may be used to exchange data between different computing devices 400 which such computing devices 400 form part an interconnected computer network. Examples of a communication interface 424 can include a modem, a network interface (such as an Ethernet card), a communication port, an antenna with associated circuitry and the like. The communication interface 424 may be wired or may be wireless. Software and data transferred via the communication interface 424 are in the form of signals which can be electronic, electromagnetic, optical or other signals capable of being received by communication interface 424. These signals are provided to the communication interface via the communication path 426.

[0045] As shown in Figure 4, the computing device 400 further includes a display interface 402 which performs operations for rendering images to an associated display

430 and an audio interface 432 for performing operations for playing audio content via associated speaker(s) 434.

[0046] As used herein, the term "computer program product" may refer, in part, to removable storage unit 418, removable storage unit 422, a hard disk installed in hard disk drive 412, or a carrier wave carrying software over communication path 426 (wireless link or cable) to communication interface 424. Computer readable storage media refers to any non-transitory tangible storage medium that provides recorded instructions and/or data to the computing device 400 for execution and/or processing. Examples of such storage media include floppy disks, magnetic tape, CD-ROM, DVD, Blu-ray™ Disc, a hard disk drive, a ROM or integrated circuit, USB memory, a magneto-optical disk, or a computer readable card such as a PCMCIA card and the like, whether or not such devices are internal or external of the computing device 400. Examples of transitory or non-tangible computer readable transmission media that may also participate in the provision of software, application programs, instructions and/or data to the computing device 400 include radio or infra-red transmission channels as well as a network connection to another computer or networked device, and the Internet or Intranets including e-mail transmissions and information recorded on Websites and the like.

[0047] The computer programs (also called computer program code) are stored in main memory 408 and/or secondary memory 410. Computer programs can also be received via the communication interface 424. Such computer programs, when executed, enable the computing device 400 to perform one or more features of embodiments discussed herein. In various embodiments, the computer programs, when executed, enable the processor 404 to perform features of the above-described embodiments. Accordingly, such computer programs represent controllers of the computer system 400.

[0048] Software may be stored in a computer program product and loaded into the computing device 400 using the removable storage drive 414, the hard disk drive 412, or the interface 420. Alternatively, the computer program product may be downloaded to the computer system 400 over the communications path 426. The software, when executed by the processor 404, causes the computing device 400 to perform functions of embodiments described herein.

[0049] It is to be understood that the embodiment of Figure 4 is presented merely by way of example. Therefore, in some embodiments one or more features of the

computing device 400 may be omitted. Also, in some embodiments, one or more features of the computing device 400 may be combined together. Additionally, in some embodiments, one or more features of the computing device 400 may be split into one or more component parts.

[0050] It will be appreciated that the elements illustrated in Figure 4 function to provide means for performing the various functions and operations of the servers as described in the above embodiments.

[0051] In an implementation, a server may be generally described as a physical device comprising at least one processor and at least one memory including computer program code. The at least one memory and the computer program code are configured to, with the at least one processor, cause the physical device to perform the requisite operations.

[0052] It will be appreciated by a person skilled in the art that numerous variations and/or modifications may be made to the present specification as shown in the specific embodiments without departing from the scope of the specification as broadly described. For example, the multi-label classification can be adapted to work with official document designs with different numbers of security features. The present embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.

CLAIMS

1. A method of determining authenticity of an official document, the method comprising:

extracting image data of a plurality of security features of the official document from an image of the official document;

applying a trained neural network to the extracted image data to determine a sharpness of each of the security features on the image of the official document; and

determining whether the official document is authentic based on a sharpness classification of the security features,

wherein training the neural network comprises:

training a multi-label classification model to classify a plurality of regions of interest, wherein the regions of interests are spatially distributed on a labeled training image, and wherein each label corresponds to a respective one of the regions of interest;

applying a spatial attention module to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest; and

optimizing parameters of the neural network based on optimizing a loss of the multi-label classification model over a plurality of labeled training images,

wherein the labeled training images comprise labeled images of official documents, and wherein the regions of interest comprise selected cropped regions containing the security features on said official documents.

2. The method as claimed in claim 1, wherein the at least one attribute comprises the sharpness of the security features on the labeled images of official documents.

3. The method as claimed in claim 1, wherein the at least one layer of the neural network comprises a last fully-connected layer.

4. The method as claimed in claim 1, wherein the official document comprises an identity card, and wherein the identity card is determined to be authentic if each of the security features on the image of the official document is classified as sharp.

5. The method as claimed in claim 1, wherein extracting the image data of the plurality of security features of the official document comprises:

extracting an image area corresponding to the official document from a single image of the official document; and

cropping the extracted image area to retain a plurality of regions containing the plurality of security features.

6. A system for determining authenticity of an official document, comprising:
a processor; and
a computer-readable memory coupled to the processor and having instructions stored thereon that are executable by the processor to:

extract image data of a plurality of security features of the official document from an image of the official document;

apply a trained neural network to the extracted image data to determine a sharpness of each of the security features on the image of the official document;
and

determine whether the official document is authentic based on a sharpness classification of the security features,

wherein the trained neural network comprises:

a multi-label classification model trained to classify a plurality of regions of interest, wherein the regions of interests are spatially distributed on a labeled training image, and wherein each label corresponds to a respective one of the regions of interest;

a spatial attention module applied to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest; and

neural network parameters optimized based on an optimization of a loss of the multi-label classification over a plurality of labeled training images,

wherein the labeled training images comprise labeled images of official documents, and wherein the regions of interest comprise selected cropped regions containing the security features on said official documents.

7. The system as claimed in claim 6, wherein the at least one attribute comprises the sharpness of the security features on the labeled images of official documents.

8. The system as claimed in claim 6, wherein the at least one layer of the neural network comprises a last fully-connected layer.

9. The system as claimed in claim 6, wherein the official document comprises an identity card, and wherein the instructions are executable by the processor to determine identity card to be authentic if each of the security features on the image of the official document is classified as sharp.

10. The method as claimed in claim 6, wherein the instructions that are executable by the processor to extract the image data of the plurality of security features of the official document further comprise instructions for:

extracting an image area corresponding to the official document from a single image of the official document; and

cropping the extracted image area to retain a plurality of regions containing the plurality of security features.

11. An apparatus comprising:

an input device for receiving an image of the official document; and

a processing device for extracting image data of a plurality of security features of the official document from the image of the official document and feeding the extracted image data to a trained neural network to determine:

a sharpness of each of the security features on the image of the official document; and

an authenticity of the official document based on a sharpness classification of the security features,

wherein the trained neural network comprises:

a multi-label classification model trained to classify a plurality of regions of interest, wherein the regions of interests are spatially distributed on a labeled training image, and wherein each label corresponds to a respective one of the regions of interest;

a spatial attention module applied to at least one layer of the neural network to learn a position and at least one attribute of each of the regions of interest; and

neural network parameters optimized based on an optimization of a loss of the multi-label classification over a plurality of labeled training images,

wherein the labeled training images comprise labeled images of official documents, and wherein the regions of interest comprise selected cropped regions containing the security features on said official documents.