

SYSTEMS, METHOD AND COMPUTER READABLE MEDIUM FOR DETECTING FORGERY OF IDENTITY DOCUMENTS USING GENERATIVE ADVERSARIAL NETWORK

TECHNICAL FIELD

[0001] The present disclosure relates generally to systems, a method and a computer readable medium for detecting forgery of identity (ID) documents. Embodiments of the disclosure are related to a Generative Adversarial Network (GAN) based system, method and computer readable medium for detecting forgery of identity (ID) documents by comparing a dewarped image obtained from a trained neural network model and a dewarped valid image generated by a trained generator model of the GAN.

BACKGROUND

[0002] Identity verification is an integral requirement in several fields and industries in the present times. With the advent of technology, identity verification processes have also been made available electronically. Electronic verification of identity (ID) documents has gained acceptance and importance in all fields and industries requiring identity verification, to enable effective identification and verification of customers or users based on the information present in the identification documents, such as passports, identity cards or driving licenses. Said fields and industries include without limitation banking and financial services, education sector, insurance sector, staffing and recruiting services, on-demand economies, peer-to-peer (P2P) marketplaces, online communities, gaming sector, telecommunication services, mobility services, health care, real estate and so on.

[0003] Identity (ID) document forgery refers to illegal manipulation by un-authorized person of ID documents issued by an authorized agency such as a government agency, to gain unauthorized access or for illegal purposes by deceiving such agency. Identity (ID) document forgery techniques have evolved rapidly these days. Fake ID detection solutions include cursory inspection, biometric verification, information extraction from the two-dimensional code or chips and the presence of microprints and

holograms which are hard to be replicated without expensive printing devices. Most of the existing solutions require specialized electronic devices to capture the images of the Identity (ID) document from different angles and to capture every single detail on the Identity (ID) documents for further analysis. Besides, the operation has to be taken under a specific environment with stable illuminance, like a fixed range of color temperature. A disadvantage of said solutions is that the quality of the captured images of the Identity (ID) document are high quality images with salient details such as clear shapes and textures. This evidently poses severe difficulty and inconvenience on the user as such conditions are often not available and/ or possible, thereby resulting in an unstable performance on lower quality images suffering from blurriness, reflection or insufficient brightness. In view of the above disadvantages, human intervention is preferred in most real-life applications of identity (ID) verification.

[0004] Moreover, another disadvantage of the conventional solutions is that verification which highly relies on specialized devices is not applicable to a lot of online service platforms. Some conventional solutions employ algorithm based approaches instead, however, it is challenging for an algorithm based approach to compete with hardware devices involved in terms of the accuracy because the images are usually taken by devices at hands like mobile phones or cameras and the quality of the image varies based on different levels of blurriness, lighting conditions and intrinsic parameters of the camera such as shutter and exposure.

[0005] Owing to the fact that a plethora of tasks are performed using bulky and costly specialized electronic hardware devices in the present day and age, there is a need for a solution that is capable of overcoming the drawbacks of the conventional solutions and enables the detection of forgery of Identity (ID) documents

SUMMARY

[0006] The present disclosure relates generally to a system, method and computer readable medium for detecting forgery of identity (ID) documents. Embodiments of the disclosure are related to a Generative Adversarial Network (GAN) based system, method and computer readable medium for detecting forgery of identity (ID) documents by comparing a dewarped image obtained from a trained neural network

model and a dewarped valid image generated by a trained generator model of the GAN.

[0007] Embodiments of the present invention can be applied to online services which mostly receive a limited number of images from the service applicants, that is, wherein captured images of the user's identity (ID) documents are restricted in terms of numbers and imaging conditions.

[0008] An advantage of the embodiments of the present invention is the ability to facilitate better performance in output and improved capability of an accurate rate. Further, another advantage of the embodiments of the present invention is that it is capable of saving time in the output arising from an increased efficiency in training than the existing models, resulting in higher accuracy.

[0009] The system, method and computer readable medium for detecting forgery of identity (ID) documents disclosed in the present application are advantageous in several ways, including but not limited to providing a robust solution for verification of captured ID documents having different brightness, orientations and/ or resolutions, or being of a lower quality and suffering from blurriness, reflection and/ or insufficient brightness, overcoming the requirement of human intervention, saving costs, time, manual effort for aligning a large number of ID documents, overcoming requirements of higher hardware specifications as utilized in conventional techniques, state-of-the-art fraud detection model, saving costs, time, manual effort and other resources employed at the stage of pre-processing data, being simpler to set up and manage, reducing the processing time significantly along with higher accuracy, obtaining an improved forgery detection system by making it robust against potential fraud ID documents that can be generated using artificial intelligence technology and are unmistakably similar to the original ID documents.

[0010] In a first aspect of the present disclosure, a system for detecting forgery of an identity (ID) document using a Generative Adversarial Network (GAN) is disclosed. The system may comprise a neural network model for converting an input image of the identity (ID) document into a new dewarped image of a different angle using a two-channel mapping function, wherein the two-channel mapping function is obtained by

training the neural network model using one or more image pair comprising at least one image of a genuine identity (ID) document and at least one transformed image, wherein the transformed image is obtained by applying a random perspective transformation on a plurality of images of the genuine identity (ID) document at different angles; the Generative Adversarial Network (GAN) comprising: a generator model is trained with an input of a plurality of images of forged identity (ID) document for generating a plurality of images of generated genuine identity (ID) document, and a discriminator model is trained with an input of a first input and a second input to generate a classification result, wherein the first input is a first concatenation image pair of the plurality of images of forged identity (ID) document and the plurality of images of the generated genuine identity (ID) document, wherein the second input is a second concatenation image pair of a plurality of manually aligned images of the genuine identity (ID) document and the plurality of images of forged identity (ID) document, wherein the discriminator model evaluates whether the plurality of the images of the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar and generates the classification result based on the evaluation, wherein the generator model after training is input with the input image of the identity (ID) document to generate a generated valid image; a processing unit verifies the legitimacy of the generated valid image and the new dewarped image and generates a difference image by calculating at least one pixel-wise difference between the generated valid image and the new dewarped image; and a Support Vector Machine (SVM) model for outputting a verification result.

[0011] According to an embodiment in conjunction to the first aspect of the present disclosure, the Support Vector Machine (SVM) model comprises a binary classifier to classify the difference image as a valid or invalid identity (ID) document.

[0012] According to an embodiment in conjunction to the first aspect of the present disclosure, the classification result is updated in the generator model and the discriminator model using a backpropagation process.

[0013] According to an embodiment in conjunction to the first aspect of the present disclosure, the neural network model is a DocUnet model comprising two successive

U-Nets.

[0014] In a second aspect of the present disclosure, a system for detecting forgery of an identity (ID) document using a Generative Adversarial Network (GAN) is disclosed. The system may comprise: a neural network model for converting an input image of the identity (ID) document into a new dewarped image of a different angle using a two-channel mapping function, the Generative Adversarial Network (GAN) comprising a generator model and a discriminator model, wherein the generator model is input with an input image of the identity (ID) document to generate a generated valid image, and a processing unit for verifying the legitimacy of the generated valid image and the new dewarped image, wherein the processing unit generates a difference image by calculating at least one pixel-wise difference between the generated valid image and the new dewarped image; and a Support Vector Machine (SVM) model for outputting a verification result.

[0015] According to an embodiment in conjunction to the second aspect of the present disclosure, the Support Vector Machine (SVM) model comprises a binary classifier to classify the difference image as a valid or invalid identity (ID) document.

[0016] According to an embodiment in conjunction to the second aspect of the present disclosure, the two-channel mapping function is obtained by training the neural network model using one or more image pairs comprising at least one image of a genuine identity (ID) document and at least one transformed image.

[0017] According to an embodiment in conjunction to the second aspect of the present disclosure, the transformed image is obtained by applying a random perspective transformation on a plurality of images of the genuine identity (ID) document at different angles.

[0018] According to an embodiment in conjunction to the second aspect of the present disclosure, the generator model is trained using a plurality of images of a forged identity (ID) document as the input to generate a plurality of images of generated genuine identity (ID) document.

[0019] According to an embodiment in conjunction to the second aspect of the present disclosure, the discriminator model is trained using a first input of a first concatenation image pair of the plurality of images of forged identity document and the plurality of images of the generated genuine identity document and a second input of a second concatenation image pair of a plurality of manually aligned images of the genuine identity document and the plurality of the forged identity document.

[0020] According to an embodiment in conjunction to the second aspect of the present disclosure, the discriminator model evaluates whether the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar and generates a classification result based on the evaluation.

[0021] According to an embodiment in conjunction to the second aspect of the present disclosure, the classification result is updated in the generator model and the discriminator model using a backpropagation process.

[0022] In a third aspect of the present disclosure, a method for detecting forgery of an identity (ID) document based on a Generative Adversarial Network (GAN) is disclosed. The method may comprise: training a neural network model using one or more image pair comprising at least one image of a genuine identity (ID) document and at least one transformed image; obtaining, by the neural network model, a two-channel mapping function; converting, by the neural network model, an input image of the identity (ID) document into a new dewarped image of a different angle using the two-channel mapping function; training the Generative Adversarial Network (GAN) using a plurality of images of forged identity (ID) document and a plurality of images of genuine identity (ID) document; inputting, to a generator model of the Generative Adversarial Network (GAN), the input image of the identity (ID) document; generating, by the generator model, the genuine valid image; verifying, by a processing unit, legitimacy of the genuine valid image and the new dewarped image; and outputting, by a Support Vector Machine (SVM) model, a verification result.

[0023] According to an embodiment in conjunction to the third aspect of the present disclosure, obtaining, by the neural network model, a two-channel mapping function comprises: obtaining a transformed image by applying a random perspective

transformation on a plurality of images of the genuine identity (ID) document at different angles; and obtaining a two-channel mapping function by training the neural network model using one or more image pair comprising at least one image of a genuine identity (ID) document and at least one transformed image.

[0024] According to an embodiment in conjunction to the third aspect of the present disclosure, training the Generative Adversarial Network (GAN) comprises: training the generator model of the Generative Adversarial Network (GAN) using a plurality of forged identity (ID) document; generating, by the generator model, a plurality of images of generated genuine identity (ID) document; obtaining, by a discriminator model of the Generative Adversarial Network (GAN), a first input comprising a first concatenation image pair of the plurality of images of forged identity (ID) document and the plurality of images of the generated genuine identity (ID) document; obtaining, by the discriminator model, a second input comprising a second concatenation image pair of a plurality of manually aligned images of the genuine identity (ID) document and the plurality of the forged identity (ID) document; training the discriminator model of the Generative Adversarial Network (GAN) using the first input and the second input; evaluating, by the discriminator model, whether the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar; and generating, by the discriminator model, a classification result.

[0025] According to an embodiment in conjunction to the third aspect of the present disclosure, training the Generative Adversarial Network (GAN) comprises updating the classification result in the generator model and the discriminator model using a backpropagation process.

[0026] According to an embodiment in conjunction to the third aspect of the present disclosure, verifying the legitimacy of the genuine valid image and the new dewarped image by the processing unit comprises: receiving the genuine valid image from the generator model; receiving the new dewarped image from the neural network model; calculating at least one pixel-wise difference between the generated valid image and the new dewarped image; and generating a difference image by calculating at least one pixel-wise difference between the generated valid image and the new dewarped image.

[0027] In a fourth aspect of the present disclosure, a computer readable medium for storing a program comprising instructions that, when executed by a system, cause the system for detecting forgery of identity (ID) document based on the Generative Adversarial Network (GAN) is disclosed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] **FIG. 1** is a non-limiting schematic diagram depicting an example system **100** for detecting forgery of an identity (ID) document using a Generative Adversarial Network (GAN) **106**, in accordance with the disclosed embodiments.

[0029] **FIG. 2** is a diagrammatic representation **200** of an example neural network model **210** for detecting forgery of identity (ID) documents, in accordance with the disclosed embodiments.

[0030] **FIG. 3** is a non-limiting schematic representation **300** of an example two-channel mapping function **310** obtained from a neural network model for converting an input image **303** into a new dewarped image **304** having a different angle, in accordance with the disclosed embodiments.

[0031] **FIG. 4** is a non-limiting schematic representation **400** of an example Generative Adversarial Network (GAN) for training a generator model **404** and a discriminator model **410** in accordance with the disclosed embodiments.

[0032] **FIG. 5** is a non-limiting schematic representation **500** of an example trained generator model **520** of a Generative Adversarial Network (GAN) for converting an input image **510** having an ID document **505** into a generated valid image **530** having an ID document **525**, in accordance with the disclosed embodiments.

[0033] **FIG. 6** is a flow diagram that illustrates an example method **600** of detecting forgery of identity (ID) documents based on a Generative Adversarial Network used, in accordance with the disclosed embodiments.

[0034] FIG. 7 is a flow diagram that illustrates an example method **700** for obtaining a two-channel mapping function from a neural network model, in accordance with the disclosed embodiments.

[0035] FIG. 8 is a flow diagram that illustrates an example method **800** used to train a generator model of a Generative Adversarial Network (GAN), in accordance with the disclosed embodiments.

[0036] FIG. 9 is a flow diagram that illustrates an example method **900** verifying legitimacy of a genuine valid image and a new dewarped image by a processing unit, in accordance with the disclosed embodiments.

[0037] The summary above, as well as the following detailed description of illustrative embodiments, is better understood when read in conjunction with the appended drawings. For the purpose of illustrating the present disclosure, exemplary constructions of the disclosure are shown in the drawings. However, the disclosure is not limited to specific methods and instrumentalities disclosed herein. Moreover, those in the art will understand that the drawings are not to scale. Wherever possible, like elements have been indicated by identical numbers.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0038] The particular configurations discussed in the following description are non-limiting examples that can be varied and are cited merely to illustrate at least one embodiment and are not intended to limit the scope thereof.

[0039] The present disclosure relates to systems, a method and a computer readable medium for detecting forgery of identity (ID) documents using a Generative Adversarial Network (GAN).

[0040] Generative Adversarial Networks (GANs) are deep-learning based generative models that use two neural networks, pitting one against the other in order to generate

new, synthetic instances of data that can pass for real data. Such architectures are widely used in image generation and voice generation. One neural network, called generator model, creates data instances, while the other neural network, called discriminator model, evaluates them for authenticity that is, the discriminator determines whether each instance of data that it reviews belongs to the actual training dataset or not. The two models, the generator and discriminator, are trained together. The generator generates a plurality of images of generated genuine identity (ID) documents. The discriminator model is trained using a first input and a second input. The first input is a first concatenation image pair of the plurality of images of forged identity (ID) document P_k and the plurality of images of the generated genuine identity (ID) document. The second input is a second concatenation image pair of a plurality of manually aligned images of the genuine identity (ID) document R_k and the plurality of the forged identity (ID) document P_k . The discriminator model evaluates whether the images of the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar and generates a classification result based on the evaluation. During training, the discriminator is updated to classify the real and synthesized data as well as possible, and the generator is updated to fool the discriminator as much as possible. Both the generator and discriminator are updated based on the classification results of the discriminator using backpropagation processes or techniques.

[0041] A neural network model is trained using at least one image of a genuine identity (ID) document and at least one transformed image to obtain a two-channel mapping function. A generator model of the Generative Adversarial Network (GAN) is trained to generate a plurality of valid images of the genuine identity (ID) document. The trained neural network model converts an input image of the identity (ID) document to be verified into a new dewarped image of different angle using the two-channel mapping function. The trained generator model generates a valid image, referred to as a generated valid image, when input with the input image. A processing unit generates a difference image by verifying the legitimacy of the generated valid image and the new dewarped image. A support vector machine classifies the difference image as a valid or invalid identity (ID) document.

[0042] FIG. 1 is a non-limiting schematic diagram depicting an example system **100**

for detecting forgery of an identity (ID) document using a Generative Adversarial Network (GAN) **106**, in accordance with the disclosed embodiments. The system **100** comprises a neural network model **104** for converting an input image **102** of the identity (ID) document to be verified into a new dewarped image **103** of different angles using a two-channel mapping function. The two-channel mapping function is obtained by training the neural network model **104** using one or more image pairs comprising at least one image of a genuine identity (ID) document and at least one transformed image. The transformed image is obtained by applying a random perspective transformation on a plurality of images of the genuine identity (ID) document at different angles. The system also uses the Generative Adversarial Network (GAN) **106** in the process of verifying the input identity (ID) document.

[0043] In an embodiment of the present invention, the Generative Adversarial Network (GAN) **106** has a generator model and a discriminator model. The generator model is trained using a plurality of images of forged identity (ID) documents as the input to generate a plurality of images of generated genuine identity (ID) documents. The discriminator model is trained using a first input and a second input. The first input is a first concatenation image pair of the plurality of images of forged identity (ID) document and the plurality of images of the generated genuine identity (ID) document. The second input is a second concatenation image pair of a plurality of manually aligned images of the genuine identity (ID) document and the plurality of the forged identity (ID) document. The discriminator model evaluates whether the images of the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar and generates a classification result based on the evaluation process. The weights of the generator model and the discriminator model are updated based on the classification result. The generator model after training using the Generative Adversarial Network (GAN) **106** is input with the input image **102** of the identity (ID) document to generate a valid image **105**.

[0044] The system **100** also comprises a processing unit **108** to verify the legitimacy of the generated valid image **105** and the new dewarped image **103** and to generate a difference image **109** by calculating at least one pixel-wise difference between the generated valid image **105** and the new dewarped image **103**. The system **100** further

comprises a support vector machine (SVM) model **110** to output a verification result. In an embodiment of the present invention, the support vector machine (SVM) model **110** comprises a binary classifier **111** to classify the difference image as a valid or invalid identity (ID) document and outputs a verification result **112** to a user.

[0045] FIG. 2 is a diagrammatic representation **200** of an example neural network model **210** for detecting forgery of identity (ID) documents, in accordance with the disclosed embodiments. In some embodiments of the present invention, said neural network model is a convolutional neural network model such as U-net. U-Net is an end-to-end fully convolutional network commonly used in image segmentation with the advantage of high accuracy and computation efficiency and is referred so due to its symmetric shape, which is different from other Fully Convolutional Network (FCN) variants. U-net is an end-to-end fully-convolutional network containing a contraction part to extract the image features followed by an expansion path for object localization. Convolutional neural networks such as U-net are preferred as they don't involve any dense layer in the architecture, thereby reducing the number of parameters and computation time. In an embodiment of the present invention, the neural network model is a DocUnet model which is composed of two successive U-nets. The example neural network model **210** disclosed in **FIG. 2** comprises two successive U-Nets **212**, **214**. In **FIG. 2**, the neural network model **210** is trained with 16000 image pairs based on 1600 flatten scanned images **122**, **124**, **126** and the corresponding images **203**, **205**, **207** taken by an imaging unit from 10 different angles. The images **203**, **205**, **207** have the identity (ID) document **202**, **204**, **206** respectively, in it. Similarly, the scanned ID images **122**, **124**, **126** have the identity (ID) document **123**, **125**, **127** respectively, in it. The scanned images **122**, **124**, **126** are obtained by taking 1600 genuine ID images from different devices, including without limitation, printers and cameras. In an example, 1600 images of real Identity (ID) documents are collected. The images may be collected by either scanning Identity (ID) documents using flatbed scanners or may be captured using other devices such as a digital camera. In an example, for each image, a plurality of corresponding images of different angles may be generated by applying a random perspective transformation, for example, ten corresponding images may be generated. In the above example, a total of 16,000 image pairs are used to train the model and each pair contains an original image **203**, **205**, **207** and a related transformed image **122**, **124**, **126**. It should be noted that the flatten scanned ID image,

the scanned ID image or the transformed image, all refer to the same and referred using the same reference numeral **122**, **124**, **126**. After training the neural network **210**, a mapping function $F(x, y)$ **126** is obtained and is used as the mapping function F **126** for input image to be verified.

[0046] **FIG. 3** is a non-limiting schematic representation **300** of an example two-channel mapping function **310** obtained from a neural network model for converting an input image **303** into a new dewarped image **304** having a different angle, in accordance with the disclosed embodiments. After the neural network model is trained in accordance with the disclosed embodiments, for example, disclosed as **104** and **210** in **FIG. 1** and **FIG. 2**, respectively, the input image I **303** is input to such neural network model and the output is a two-channel mapping function $F(x, y)$ **310**, where (x, y) is the coordinates of image I **303** and $F(x, y)$ **310** denotes the corresponding coordinates in the dewarped image I_D **304**. The input image I **303** refers to the image to be verified sent by an applicant and also referred to as the original image. The mapping function F **310** converts the input image I **303** to a dewarped image I_D **304** of a different angle. The neural network model dewarps the input image I **303** with identity (ID) document **302** via the mapping function F **310**. The input image I **303** is flattened so that the result image is closer to the image taken from the top. The dewarped image I_D **304** with ID document **305** preserves the image features including color, brightness and all the low and high frequency information. After rectifying the input image I **303**, the image is dewarped while the image features are preserved in the ID document. The rectified or dewarped image I_D **304** is obtained by applying the mapping function F **310** to input image I **303**. The rectified image and dewarped image I_D **304** both are referred to as the output of the trained neural network model.

[0047] **FIG. 4** is a non-limiting schematic representation **400** of an example Generative Adversarial Network (GAN) for training a generator model **404** and a discriminator model **410** in accordance with the disclosed embodiments. The Generative Adversarial Network (GAN) has a generator model **404** and a discriminator model **410**. The generator model **404** is trained using a plurality of images of forged identity (ID) document as the input P_k **402** to generate a plurality of images of generated genuine identity (ID) document **406**. The discriminator model **410** is trained

using a first input **407** and a second input **409**. The first input **407** is a first concatenation image pair of the plurality of images of forged identity (ID) document P_k **402** and the plurality of images of the generated genuine identity (ID) document **406**. The second input **409** is a second concatenation image pair of a plurality of manually aligned images of the genuine identity (ID) document R_k **408** and the plurality of the forged identity (ID) document P_k **402**. The discriminator model **410** evaluates whether the images of the generated genuine identity (ID) document **406** and the manually aligned image of the genuine identity (ID) document R_k **408** are similar and generates a classification result **412** based on the evaluation. The weights of the generator model **404** and the discriminator model **410** are updated based on the classification result **412**. The generator model **404** after training is input with the input image of the identity (ID) document to generate a valid image, for example, the valid image **105**, as disclosed in **FIG. 1** of the present disclosure.

[0048] In an embodiment of the present invention, Generative Adversarial Network (GAN) is applied to generate images which are regarded as the valid Identity (ID) documents of the potential fake Identity (ID) documents. Generative adversarial network (GAN) is one of the most popular techniques in computer vision. It is commonly used to generate new examples that plausibly come from an existing distribution of samples, such as generating new photographs that are similar but specifically different from a dataset of existing photographs. In general, the generative model can be used to generate new plausible data on demand such as aging faces, realistic images from sketches or higher resolution images.

[0049] In general, the Generative Adversarial Network (GAN) consists of two neural network models. The generator model learns to generate new plausible samples to fool the second model called the discriminator model that learns to differentiate generated examples from real examples. In an embodiment of the present invention, pix2pixHD, which is a well-known image-to-image transformation framework based on conditional settings, may be employed to render a genuine ID image I_G . The pix2pixHD framework is known for rendering photorealistic images at 2048 x 1024 resolution. In an embodiment of the present invention, the generator model is trained with a plurality of valid ID images, for example 1,600 valid images and a plurality of images of corresponding fake Identity (ID) documents, for example 7,500 images of fake Identity

(ID) documents. It is to be noted that the images of fake Identity (ID) documents are preferred to cover various illuminance conditions including different levels of shadow, glare and brightness as well as multiple perspective views. In the training stage, the generator model takes a forged ID image P_k **402** as the input and generates the plurality of images of generated genuine identity (ID) document **406**. Two image pairs are fed as the input of the discriminator model **410**. The first pair or first input **407** is the concatenation of the input image P_k **402** and the image of the generated genuine identity (ID) document **406** from the generator model **404**. The second pair or first input **409** is the concatenation of the input image P_k **402** and the manually aligned image of the genuine identity (ID) document R_k **408**. The discriminator loss is minimized to better differentiate the generated pair from the target pair and the generator loss is minimized for the purpose of generating a more photorealistic genuine ID image. In an embodiment of the present invention, the generator model **404** is only used for verifying the input image to be verified and the discriminator model **410** is not involved in the inference stage.

[0050] In general, the Generative Adversarial Network (GAN) is used to generate fake or non-existing image from the real samples while the present invention uses Generative Adversarial Network (GAN) to train the generator model **404** to generate a genuine ID images **406** from the forged ID images **402**. In an embodiment of the present invention, a set of image pairs (P_k , R_k) **402**, **408** are used for training the Generative Adversarial Network (GAN), where R_k **408** is a plurality of manually aligned images of the genuine identity (ID) document and P_k **402** is a combination of two sets, for example, images of different perspective and environment conditions and images with forgeries. P_k **402** is the input of the generator model **404** and the concatenation of P_k **402** and R_k **408** is the input of discriminator model **410**. Another input of the discriminator model **410** is the pair of P_k **402** and the ID image **406** generated by the generator model **404**. The discriminator model **410** evaluates whether the generated image **406** is similar to the genuine image R_k **408** and provides a classification result **412** based on the two inputs and updates the weights of generator model **404** and discriminator model **410** accordingly. The losses or inaccuracies of the discriminator model **410** and the generator model **404**, which are indicative of the errors in output, are determined, and the weights for the two models are updated, respectively, using a feedback loop or backpropagation process **413**. Backpropagation process **413** refers

to a backward propagation of error in a learning of neural networks using gradient descent. In the backpropagation process **413**, gradient descent is used to iteratively find a set of parameters or weights for the models that minimises the loss or error. Gradient descent involves using the derivative of a function to adjust the value of a parameter in an attempt to bring it to convergence. Accordingly, by way of backpropagation **413**, the losses or errors adjust the weights of the generator model **404** so that the generator model **404** can generate more and more realistic images of generated genuine identity (ID) document **406** as it is trained further. Further, by way of backpropagation process **413**, the losses or errors will adjust the weights of the discriminator model **410** so that said model **410** is able to more effectively verify the authenticity of the first input **407** and a second input **409**. In a context of a neural network, back-propagation process **413** would happen for the weights associated with each neuron in the network, and each iteration of training the neural network will adjust these weights towards minimizing the loss or error of the network. In an embodiment of the present invention, the generator model **404** and the discriminator model **410** are trained until the losses are minimized through the backpropagation process (feedback loop) **413** as disclosed in **FIG. 4**.

[0051] FIG. 5 is a non-limiting schematic representation **500** of an example trained generator model **520** of a Generative Adversarial Network (GAN) for converting an input image **510** having an ID document **505** into a generated valid image **530** having an ID document **525**, in accordance with the disclosed embodiments. Genuine ID images generated by a generator model of the Generative Adversarial Network (GAN) have characteristic such as, the images are photorealistic images with high image resolution and dynamic range, the illuminance conditions such as glare and shadow are alleviated and the potential forged part on the ID is corrected. In an embodiment of the present invention, the input image **510** is input to the generator model **520** and a neural network model. The generated valid image or genuine ID image obtained from the generator model and a neural network model are the same, however, the generator model of the Generative Adversarial Network (GAN) corrects some artefacts, such as shadow and increases the image resolution. In **FIG. 5**, the Identity (ID) document **525** in image I_G **530** represents consistent features including brightness, resolution and free of artefacts. The generated Identity (ID) document I_G is taken as the certified document and compared with the unwrapped Identity (ID)

document resulting from the neural network model.

[0052] The neural network model is applied to flatten and unwarp the input image and keep all the details including brightness, textures and dynamic range. The Generative Adversarial Network (GAN), for example pix2pixHD model, on the other hand, renders a high resolution image without potential forgeries and extreme color distribution suffering from lighting conditions. In an embodiment of the present invention, a support vector machine (SVM) model classifies the difference of the two images (generated by the neural network model and the Generative Adversarial Network (GAN)) and outputs a verification result which indicates if the input image is genuine.

[0053] **FIG. 6** is a flow diagram that illustrates an example method **600** of detecting forgery of identity (ID) documents based on a Generative Adversarial Network used, in accordance with the disclosed embodiments. As at block **602**, a neural network model is trained using one or more image pairs comprising at least one image of a genuine identity (ID) document and at least one transformed image. In some embodiments of the present invention, said neural network model is a convolutional neural network model such as U-net. U-net is one of the well-known and commonly employed convolutional neural networks being a Fully Convolutional Network (FCN) that does image segmentation and is referred so due to its symmetric shape, which is different from other FCN variants. U-net is an end-to-end fully-convolutional network containing a contraction part to extract the image features followed by an expansion path for object localization. Convolutional neural networks such as U-net are preferred as they don't involve any dense layer in the architecture, thereby reducing the number of parameters and computation time. In an embodiment of the present invention, the neural network model is a DocUnet model which is composed of two successive U-nets. As at block **604**, the neural network model obtains a two-channel mapping function after training the neural network model. Then, as at block **606**, the neural network model converts an input image of the identity (ID) document to be verified into a new dewarped image of different angle using a two-channel mapping function. As at block **608**, the Generative Adversarial Network (GAN) is sufficiently trained using a plurality of images of forged identity (ID) document and a plurality of images of genuine

identity (ID) document, such that a generator model of the generative adversarial network generates a genuine valid image. Then as at block **610**, the generator model after training is input with the input image of the identity (ID) document to be verified.

[0054] As at block **612**, the genuine valid image is obtained from the generator model. As at block **614**, a processing unit verifies the legitimacy of the genuine valid image from the generator model and the new dewarped image from the neural network model. Then, the processing unit generates a difference image by calculating at least one pixel-wise difference between the generated valid image and the new dewarped image, as at block **616**. A support vector machine classifies the difference image as a valid or invalid identity (ID) document and outputs a verification result to the user, as at block **618**.

[0055] **FIG. 7** is a flow diagram that illustrates an example method **700** for obtaining a two-channel mapping function from a neural network model, in accordance with the disclosed embodiments. As at block **702**, a transformed image is obtained by applying a random perspective transformation on a plurality of images of a genuine identity (ID) document at different angles. Then as at block **704**, the two-channel mapping function is obtained by training the neural network model using one or more image pairs comprising at least one image of a genuine identity (ID) document and at least one transformed image.

[0056] **FIG. 8** is a flow diagram that illustrates an example method **800** used to train a generator model of a Generative Adversarial Network (GAN), in accordance with the disclosed embodiments. As at block **802**, the generator model of the Generative Adversarial Network (GAN) is trained using a plurality of forged identity (ID) documents. As at block **804**, the generator model generates a plurality of images of a generated genuine identity (ID) document after training. As at block **806**, a discriminator model is used to obtain a first input comprising a first concatenation image pair of the plurality of images of a forged identity (ID) document and the plurality of images of the generated genuine identity (ID) document. As at block **808**, the discriminator model is used to obtain a second input comprising a second concatenation image pair of a plurality of manually aligned images of the genuine

identity (ID) document and the plurality of the forged identity (ID) document.

[0057] The discriminator model of the Generative Adversarial Network (GAN) is trained using the first input and the second input, as at block **810**. The discriminator model evaluates whether the plurality of images of the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar and generates a classification result, as at the block **812**. Finally, as at block **814**, weights of the generator model and the discriminator model are updated based on the classification result. By updating the weights of the general model and the discriminator model (also referred to as the backpropagation process or feedback loops in the system), both models are able to improve their efficiency based on the classification determined by the discriminator model in respect of an input image synthesized by the generator model.

[0058] **FIG. 9** is a flow diagram that illustrates an example method **900** verifying legitimacy of a genuine valid image and a new dewarped image by a processing unit, in accordance with the disclosed embodiments. As at block **902**, a support vector machine (SVM) model receives the genuine valid image from a generator model. The support vector machine (SVM) model receives the new dewarped image from a neural network model, as at block **904**. The support vector machine (SVM) model calculates at least one pixel-wise difference between the generated valid image and the new dewarped image and generates a difference image from the pixel-wise difference, as at block **906** and **908**.

[0059] For example, the support vector machine (SVM) model calculates the pixel-wise difference between a generated valid ID image I_G and de-skewed (new dewarped) image I_D to verify the legitimacy. The pixel value shows the difference of I_G and I_D . The higher pixel values explicitly indicate the significant illuminance difference and the potential manipulation. The images I_G and I_D are both unwrapped versions of image I . The forgery and artefact due to environmental issues while taking the picture in image I is alleviated in I_G . The present invention distinguishes between the two cases, that is, of forgery and artefact due to environmental issues as it is pertinent to identify and classify cases of forgery and not environment artefacts. The support

vector machine (SVM) model performs an important role in identifying cases of real forgery. The support vector machine (SVM) model, a supervised machine learning approach which not only performs well in high dimensional data but is also memory efficient and works as a binary classifier to categorize the difference image into two classes. In the training process, the support vector machine (SVM) model learns to identify if there exists any forgery on the card surface and the high values in the feature come from fabricated areas are recognized while those resulting from illuminance variance are omitted. Features with higher scores in the first class shows the input image is a real ID and those falls in the second class indicate the input image is invalid.

[0060] In an embodiment of the present invention, the training set of support vector machine (SVM) model is given as a set of difference images and each difference image is the absolute difference of a pair of corresponding images $\{(r_i, f_{i,k})\}$, where r_i is an image in real ID set with 1600 images and $f_{i,k}$ is the corresponding image with defect. It is to be noted that both r_i and $f_{i,k}$ are manually aligned so they share the same layout. Each difference image is considered as a data sample associated with a two-class label. The label indicates whether the data sample comes from a forged ID or not.

[0061] The present invention is a software based approach that overcomes the requirement of a specific hardware device. Embodiments of the present invention can be applied to any online platform or mobile applications which require ID validation or verification. As disclosed above, embodiments of the present invention can be applied in fields and industries, including without limitation, banking and financial services, education sector, insurance sector, staffing and recruiting services, on-demand economies, peer-to-peer (P2P) marketplaces, online communities, gaming sector, telecommunication services, mobility services, health care, real estate and so on. The techniques of forging Identity (ID) documents have evolved in the present times, hence the subtle artefacts are usually imperceptible by human perception. The present invention relies on neural networks trained with enormous fake ID images and real images which can discover the potential defect of the fake Identity (ID) documents.

[0062] The present invention uses Generative Adversarial Networks (GANs) to

generate valid ID images which are illuminance consistent, properly aligned, and fixed dynamic ranges. Embodiments of the present invention have a significantly decreased false rejection and acceptance rate owing to the fact that the generated valid ID provides a consistent basis to calculate the difference image which highly increases the accuracy of the classifier in the support vector machine (SVM) model.

[0063] It is an object of the present invention to provide systems, a method and a computer readable medium having a fast processing time, preferably, the processing time for each image being 2.5 to 3 seconds on average. It is another object of the present invention to provide systems, a method and a computer readable medium to be acceptable for online real-time validation processes and being much more efficient than validation operations done with electronic devices since it includes some device setup and calibration steps. An advantage of the present invention is that the genuine ID generated by Generative Adversarial Network (GAN) has higher image quality, standard dynamic range and does not suffer from illuminance issues like low brightness and glare. In an embodiment of the present invention, the intermediate result of the present invention may be integrated with other software based solutions to further verify all the information on the cards such as biometrics, barcode and personal profile against the central database.

[0064] It will be appreciated that variations of the above disclosed and other features and functions, or alternatives thereof, may be desirably combined into many other different systems or applications. Also, various presently unforeseen or unanticipated alternatives, modifications, variations or improvements therein may be subsequently made by those skilled in the art which are also intended to be encompassed by the following claims.

[0065] Although embodiments of the current disclosure have been described comprehensively in considerable detail to cover the possible aspects, those skilled in the art would recognize that other versions of the disclosure are also possible.

CLAIMS

What is claimed is:

1. A system for detecting forgery of an identity (ID) document using a Generative Adversarial Network (GAN), the system comprising:

a neural network model for converting an input image of the identity (ID) document into a new dewarped image of a different angle using a two-channel mapping function,

wherein the two-channel mapping function is obtained by training the neural network model using one or more image pair comprising at least one image of a genuine identity (ID) document and at least one transformed image, wherein the transformed image is obtained by applying a random perspective transformation on a plurality of images of the genuine identity (ID) document at different angles;

the Generative Adversarial Network (GAN) comprising:

a generator model is trained with an input of a plurality of images of forged identity (ID) document for generating a plurality of images of generated genuine identity (ID) document, and

a discriminator model is trained with an input of a first input and a second input to generate a classification result,

wherein the first input is a first concatenation image pair of the plurality of images of forged identity (ID) document and the plurality of images of the generated genuine identity (ID) document,

wherein the second input is a second concatenation image pair of a plurality of manually aligned images of the genuine identity (ID) document and the plurality of images of forged identity (ID) document,

wherein the discriminator model evaluates whether the plurality of the images of the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar and generates the

classification result based on the evaluation,

wherein the generator model after training is input with the input image of the identity (ID) document to generate a generated valid image;

a processing unit verifies the legitimacy of the generated valid image and the new dewarped image and generates a difference image by calculating at least one pixel-wise difference between the generated valid image and the new dewarped image; and

a Support Vector Machine (SVM) model for outputting a verification result.

2. A system for detecting forgery of an identity (ID) document using a Generative Adversarial Network (GAN), the system comprising:

a neural network model for converting an input image of the identity (ID) document into a new dewarped image of a different angle using a two-channel mapping function,

the Generative Adversarial Network (GAN) comprising a generator model and a discriminator model,

wherein the generator model is input with an input image of the identity (ID) document to generate a generated valid image, and

a processing unit for verifying the legitimacy of the generated valid image and the new dewarped image

wherein the processing unit generates a difference image by calculating at least one pixel-wise difference between the generated valid image and the new dewarped image; and

a Support Vector Machine (SVM) model for outputting a verification result.

3. The system of any one of claims **1** or **2**, wherein the Support Vector Machine (SVM) model comprises a binary classifier to classify the difference image as a valid or invalid identity (ID) document.

4. The system of claim **2**, wherein the two-channel mapping function is obtained by training the neural network model using one or more image pairs comprising at least one image of a genuine identity (ID) document and at least

one transformed image.

5. The system of claim 4, wherein the transformed image is obtained by applying a random perspective transformation on a plurality of images of the genuine identity (ID) document at different angles.
6. The system of claim 2, wherein the generator model is trained using a plurality of images of a forged identity (ID) document as the input to generate a plurality of images of generated genuine identity (ID) document.
7. The system of claim 2, wherein the discriminator model is trained using a first input of a first concatenation image pair of the plurality of images of forged identity document and the plurality of images of the generated genuine identity document and a second input of a second concatenation image pair of a plurality of manually aligned images of the genuine identity document and the plurality of the forged identity document
8. The system of claim 7, wherein the discriminator model evaluates whether the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar and generates a classification result based on the evaluation.
9. The system of any one of claims 1 or 8, wherein the classification result is updated in the generator model and the discriminator model using a backpropagation process.
10. The system of any one of claims 1 or 2, wherein the neural network model is a DocUnet model comprising two successive U-Nets.
11. A method for detecting forgery of an identity (ID) document based on a Generative Adversarial Network (GAN), comprising:
 training a neural network model using one or more image pair comprising at least one image of a genuine identity (ID) document and at least one

transformed image;

obtaining, by the neural network model, a two-channel mapping function;

converting, by the neural network model, an input image of the identity (ID) document into a new dewarped image of a different angle using the two-channel mapping function;

training the Generative Adversarial Network (GAN) using a plurality of images of forged identity (ID) document and a plurality of images of genuine identity (ID) document;

inputting, to a generator model of the Generative Adversarial Network (GAN), the input image of the identity (ID) document;

generating, by the generator model, the genuine valid image;

verifying, by a processing unit, legitimacy of the genuine valid image and the new dewarped image; and

outputting, by a Support Vector Machine (SVM) model, a verification result.

12. The method of claim **11**, wherein obtaining, by the neural network model, a two-channel mapping function comprises:

obtaining a transformed image by applying a random perspective transformation on a plurality of images of the genuine identity (ID) document at different angles; and

obtaining a two-channel mapping function by training the neural network model using one or more image pair comprising at least one image of a genuine identity (ID) document and at least one transformed image.

13. The method of claim **11**, wherein training the Generative Adversarial Network (GAN) comprises:

training the generator model of the Generative Adversarial Network (GAN) using a plurality of forged identity (ID) document;

generating, by the generator model, a plurality of images of generated genuine identity (ID) document;

obtaining, by a discriminator model of the Generative Adversarial Network (GAN), a first input comprising a first concatenation image pair of the plurality of images of forged identity (ID) document and the plurality of images

of the generated genuine identity (ID) document;

obtaining, by the discriminator model, a second input comprising a second concatenation image pair of a plurality of manually aligned images of the genuine identity (ID) document and the plurality of the forged identity (ID) document;

training the discriminator model of the Generative Adversarial Network (GAN) using the first input and the second input;

evaluating, by the discriminator model, whether the generated genuine identity (ID) document and manually aligned images of the genuine identity (ID) document are similar; and

generating, by the discriminator model, a classification result.

14. The method of claim **13**, wherein training the Generative Adversarial Network (GAN) comprises updating the classification result in the generator model and the discriminator model using a backpropagation process.

15. The method of claim **11**, wherein verifying the legitimacy of the genuine valid image and the new dewarped image by the processing unit comprises:

receiving the genuine valid image from the generator model;

receiving the new dewarped image from the neural network model;

calculating at least one pixel-wise difference between the generated valid image and the new dewarped image; and

generating a difference image by calculating at least one pixel-wise difference between the generated valid image and the new dewarped image.

16. A computer readable medium for storing a program comprising instructions that, when executed by a system, cause the system for detecting forgery of identity (ID) document based on the Generative Adversarial Network (GAN), of any of claims **11-15**.